6 Orthogonality and Least Squares

6.1 SOLUTIONS

Notes: The first half of this section is computational and is easily learned. The second half concerns the concepts of orthogonality and orthogonal complements, which are essential for later work. Theorem 3 is an important general fact, but is needed only for Supplementary Exercise 13 at the end of the chapter and in Section 7.4. The optional material on angles is not used later. Exercises 27–31 concern facts used later.

1. Since $\mathbf{u}=\begin{bmatrix}-1\\2\end{bmatrix}$ and $\mathbf{v}=\begin{bmatrix}4\\6\end{bmatrix}$, $\mathbf{u}\cdot\mathbf{u}=(-1)^2+2^2=5$, $\mathbf{v}\cdot\mathbf{u}=4(-1)+6(2)=8$, and $\dfrac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}=\dfrac{8}{5}$.

2. Since $\mathbf{w}=\begin{bmatrix}3\\-1\\-5\end{bmatrix}$ and $\mathbf{x}=\begin{bmatrix}6\\-2\\3\end{bmatrix}$, $\mathbf{w}\cdot\mathbf{w}=3^2+(-1)^2+(-5)^2=35$, $\mathbf{x}\cdot\mathbf{w}=6(3)+(-2)(-1)+3(-5)=5$,

and $\dfrac{\mathbf{x}\cdot\mathbf{w}}{\mathbf{w}\cdot\mathbf{w}}=\dfrac{5}{35}=\dfrac{1}{7}$.

3. Since $\mathbf{w}=\begin{bmatrix}3\\-1\\-5\end{bmatrix}$, $\mathbf{w}\cdot\mathbf{w}=3^2+(-1)^2+(-5)^2=35$, and $\dfrac{1}{\mathbf{w}\cdot\mathbf{w}}\mathbf{w}=\begin{bmatrix}3/35\\-1/35\\-1/7\end{bmatrix}$.

4. Since $\mathbf{u}=\begin{bmatrix}-1\\2\end{bmatrix}$, $\mathbf{u}\cdot\mathbf{u}=(-1)^2+2^2=5$ and $\dfrac{1}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}=\begin{bmatrix}-1/5\\2/5\end{bmatrix}$.

5. Since $\mathbf{u}=\begin{bmatrix}-1\\2\end{bmatrix}$ and $\mathbf{v}=\begin{bmatrix}4\\6\end{bmatrix}$, $\mathbf{u}\cdot\mathbf{v}=(-1)(4)+2(6)=8$, $\mathbf{v}\cdot\mathbf{v}=4^2+6^2=52$, and

$$\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}=\frac{2}{13}\begin{bmatrix}4\\6\end{bmatrix}=\begin{bmatrix}8/13\\12/13\end{bmatrix}.$$

6. Since $\mathbf{x}=\begin{bmatrix}6\\-2\\3\end{bmatrix}$ and $\mathbf{w}=\begin{bmatrix}3\\-1\\-5\end{bmatrix}$, $\mathbf{x}\cdot\mathbf{w} = 6(3) + (-2)(-1) + 3(-5) = 5$, $\mathbf{x}\cdot\mathbf{x}=6^2+(-2)^2+3^2=49$, and

$$\left(\frac{\mathbf{x}\cdot\mathbf{w}}{\mathbf{x}\cdot\mathbf{x}}\right)\mathbf{x}=\frac{5}{49}\begin{bmatrix}6\\-2\\3\end{bmatrix}=\begin{bmatrix}30/49\\-10/49\\15/49\end{bmatrix}.$$

7. Since $\mathbf{w}=\begin{bmatrix}3\\-1\\-5\end{bmatrix}$, $\|\mathbf{w}\|=\sqrt{\mathbf{w}\cdot\mathbf{w}}=\sqrt{3^2+(-1)^2+(-5)^2}=\sqrt{35}.$

8. Since $\mathbf{x}=\begin{bmatrix}6\\-2\\3\end{bmatrix}$, $\|\mathbf{x}\|=\sqrt{\mathbf{x}\cdot\mathbf{x}}=\sqrt{6^2+(-2)^2+3^2}=\sqrt{49}=7.$

9. A unit vector in the direction of the given vector is

$$\frac{1}{\sqrt{(-30)^2+40^2}}\begin{bmatrix}-30\\40\end{bmatrix}=\frac{1}{50}\begin{bmatrix}-30\\40\end{bmatrix}=\begin{bmatrix}-3/5\\4/5\end{bmatrix}$$

10. A unit vector in the direction of the given vector is

$$\frac{1}{\sqrt{(-6)^2+4^2+(-3)^2}}\begin{bmatrix}-6\\4\\-3\end{bmatrix}=\frac{1}{\sqrt{61}}\begin{bmatrix}-6\\4\\-3\end{bmatrix}=\begin{bmatrix}-6/\sqrt{61}\\4/\sqrt{61}\\-3\sqrt{61}\end{bmatrix}$$

11. A unit vector in the direction of the given vector is

$$\frac{1}{\sqrt{(7/4)^2+(1/2)^2+1^2}}\begin{bmatrix}7/4\\1/2\\1\end{bmatrix}=\frac{1}{\sqrt{69/16}}\begin{bmatrix}7/4\\1/2\\1\end{bmatrix}=\begin{bmatrix}7/\sqrt{69}\\2/\sqrt{69}\\4/\sqrt{69}\end{bmatrix}$$

12. A unit vector in the direction of the given vector is

$$\frac{1}{\sqrt{(8/3)^2+2^2}}\begin{bmatrix}8/3\\2\end{bmatrix}=\frac{1}{\sqrt{100/9}}\begin{bmatrix}8/3\\2\end{bmatrix}=\begin{bmatrix}4/5\\3/5\end{bmatrix}$$

13. Since $\mathbf{x}=\begin{bmatrix}10\\-3\end{bmatrix}$ and $\mathbf{y}=\begin{bmatrix}-1\\-5\end{bmatrix}$, $\|\mathbf{x}-\mathbf{y}\|^2=[10-(-1)]^2+[-3-(-5)]^2=125$ and $\text{dist}\,(\mathbf{x},\mathbf{y})=\sqrt{125}=5\sqrt{5}.$

14. Since $\mathbf{u}=\begin{bmatrix}0\\-5\\2\end{bmatrix}$ and $\mathbf{z}=\begin{bmatrix}-4\\-1\\8\end{bmatrix}$, $\|\mathbf{u}-\mathbf{z}\|^2=[0-(-4)]^2+[-5-(-1)]^2+[2-8]^2=68$ and dist $(\mathbf{u},\mathbf{z})=\sqrt{68}=2\sqrt{17}$.

15. Since $\mathbf{a}\cdot\mathbf{b}=8(-2)+(-5)(-3)=-1\neq 0$, **a** and **b** are not orthogonal.

16. Since $\mathbf{u}\cdot\mathbf{v}=12(2)+(3)(-3)+(-5)(3)=0$, **u** and **v** are orthogonal.

17. Since $\mathbf{u}\cdot\mathbf{v}=3(-4)+2(1)+(-5)(-2)+0(6)=0$, **u** and **v** are orthogonal.

18. Since $\mathbf{y}\cdot\mathbf{z}=(-3)(1)+7(-8)+4(15)+0(-7)=1\neq 0$, **y** and **z** are not orthogonal.

19. **a**. True. See the definition of $\|\mathbf{v}\|$.
b. True. See Theorem 1(c).
c. True. See the discussion of Figure 5.
d. False. Counterexample: $\begin{bmatrix}1&1\\0&0\end{bmatrix}$.
e. True. See the box following Example 6.

20. **a**. True. See Example 1 and Theorem 1(a).
b. False. The absolute value sign is missing. See the box before Example 2.
c. True. See the defintion of orthogonal complement.
d. True. See the Pythagorean Theorem.
e. True. See Theorem 3.

21. Theorem 1(b): $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w}=(\mathbf{u}+\mathbf{v})^T\mathbf{w}=(\mathbf{u}^T+\mathbf{v}^T)\mathbf{w}=\mathbf{u}^T\mathbf{w}+\mathbf{v}^T\mathbf{w}=\mathbf{u}\cdot\mathbf{w}+\mathbf{v}\cdot\mathbf{w}$. The second and third equalities used Theorems 3(b) and 2(c), respectively, from Section 2.1. Theorem 1(c): $(c\mathbf{u})\cdot\mathbf{v}=(c\mathbf{u})^T\mathbf{v}=c(\mathbf{u}^T\mathbf{v})=c(\mathbf{u}\cdot\mathbf{v})$. The second equality used Theorems 3(c) and 2(d), respectively, from Section 2.1.

22. Since $\mathbf{u}\cdot\mathbf{u}$ is the sum of the squares of the entries in **u**, $\mathbf{u}\cdot\mathbf{u}\geq 0$. The sum of squares of numbers is zero if and only if all the numbers are themselves zero.

23. One computes that $\mathbf{u}\cdot\mathbf{v}=2(-7)+(-5)(-4)+(-1)6=0$, $\|\mathbf{u}\|^2=\mathbf{u}\cdot\mathbf{u}=2^2+(-5)^2+(-1)^2=30$, $\|\mathbf{v}\|^2=\mathbf{v}\cdot\mathbf{v}=(-7)^2+(-4)^2+6^2=101$, and $\|\mathbf{u}+\mathbf{v}\|^2=(\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})=$ $(2+(-7))^2+(-5+(-4))^2+(-1+6)^2=131$.

24. One computes that $\|\mathbf{u}+\mathbf{v}\|^2=(\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})=\mathbf{u}\cdot\mathbf{u}+2\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{v}=\|\mathbf{u}\|^2+2\mathbf{u}\cdot\mathbf{v}+\|\mathbf{v}\|^2$ and $\|\mathbf{u}-\mathbf{v}\|^2=(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})=\mathbf{u}\cdot\mathbf{u}-2\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{v}=\|\mathbf{u}\|^2-2\mathbf{u}\cdot\mathbf{v}+\|\mathbf{v}\|^2$, so $\|\mathbf{u}+\mathbf{v}\|^2+\|\mathbf{u}-\mathbf{v}\|^2=\|\mathbf{u}\|^2+2\mathbf{u}\cdot\mathbf{v}+\|\mathbf{v}\|^2+\|\mathbf{u}\|^2-2\mathbf{u}\cdot\mathbf{v}+\|\mathbf{v}\|^2=2\|\mathbf{u}\|^2+2\|\mathbf{v}\|^2$.

25. When $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$, the set H of all vectors $\begin{bmatrix} x \\ y \end{bmatrix}$ that are orthogonal to $\boldsymbol{v}$ is the subspace of vectors whose entries satisfy $ax + by = 0$. If $a \neq 0$, then $x = -(b/a)y$ with y a free variable, and H is a line through the origin. A natural choice for a basis for H in this case is $\left\{ \begin{bmatrix} -b \\ a \end{bmatrix} \right\}$. If $a = 0$ and $b \neq 0$, then $by = 0$. Since $b \neq 0$, $y = 0$ and x is a free variable. The subspace H is again a line through the origin. A natural choice for a basis for H in this case is $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$, but $\left\{ \begin{bmatrix} -b \\ a \end{bmatrix} \right\}$ is still a basis for H since $a = 0$ and $b \neq 0$. If $a = 0$ and $b = 0$, then $H = \mathbb{R}^2$ since the equation $0x + 0y = 0$ places no restrictions on x or y.

26. Theorem 2 in Chapter 4 may be used to show that W is a subspace of $\mathbb{R}^3$, because W is the null space of the 1×3 matrix $\mathbf{u}^T$. Geometrically, W is a plane through the origin.

27. If $\mathbf{y}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$, then $\mathbf{y} \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{v} = 0$, and hence by a property of the inner product, $\mathbf{y} \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{y} \cdot \mathbf{u} + \mathbf{y} \cdot \mathbf{v} = 0 + 0 = 0$. Thus $\mathbf{y}$ is orthogonal to $\mathbf{u} + \mathbf{v}$.

28. An arbitrary $\mathbf{w}$ in Span$\{\mathbf{u}, \mathbf{v}\}$ has the form $\mathbf{w} = c_1\mathbf{u} + c_2\mathbf{v}$. If $\mathbf{y}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$, then $\mathbf{u} \cdot \mathbf{y} = \mathbf{v} \cdot \mathbf{y} = 0$. By Theorem 1(b) and 1(c),

$$\mathbf{w} \cdot \mathbf{y} = (c_1\mathbf{u} + c_2\mathbf{v}) \cdot \mathbf{y} = c_1(\mathbf{u} \cdot \mathbf{y}) + c_2(\mathbf{v} \cdot \mathbf{y}) = 0 + 0 = 0$$

29. A typical vector in W has the form $\mathbf{w} = c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p$. If $\mathbf{x}$ is orthogonal to each $\mathbf{v}_j$, then by Theorems 1(b) and 1(c),

$$\mathbf{w} \cdot \mathbf{x} = (c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p) \cdot \mathbf{x} = c_1(\mathbf{v}_1 \cdot \mathbf{x}) + \ldots + c_p(\mathbf{v}_p \cdot \mathbf{x}) = 0$$

So $\mathbf{x}$ is orthogonal to each $\mathbf{w}$ in W.

30. **a**. If $\mathbf{z}$ is in $W^\perp$, $\mathbf{u}$ is in W, and c is any scalar, then $(c\mathbf{z}) \cdot \mathbf{u} = c(\mathbf{z} \cdot \mathbf{u}) = c0 = 0$. Since $\mathbf{u}$ is any element of W, $c\mathbf{z}$ is in $W^\perp$.

b. Let $\mathbf{z}_1$ and $\mathbf{z}_2$ be in $W^\perp$. Then for any $\mathbf{u}$ in W, $(\mathbf{z}_1 + \mathbf{z}_2) \cdot \mathbf{u} = \mathbf{z}_1 \cdot \mathbf{u} + \mathbf{z}_2 \cdot \mathbf{u} = 0 + 0 = 0$. Thus $\mathbf{z}_1 + \mathbf{z}_2$ is in $W^\perp$.

c. Since $\mathbf{0}$ is orthogonal to every vector, $\mathbf{0}$ is in $W^\perp$. Thus $W^\perp$ is a subspace.

31. Suppose that $\mathbf{x}$ is in W and $W^\perp$. Since $\mathbf{x}$ is in $W^\perp$, $\mathbf{x}$ is orthogonal to every vector in W, including $\mathbf{x}$ itself. So $\mathbf{x} \cdot \mathbf{x} = 0$, which happens only when $\mathbf{x} = \mathbf{0}$.

32. **[M]**

a. One computes that $\|\mathbf{a}_1\| = \|\mathbf{a}_2\| = \|\mathbf{a}_3\| = \|\mathbf{a}_4\| = 1$ and that $\mathbf{a}_i \cdot \mathbf{a}_j = 0$ for $i \neq j$.

b. Answers will vary, but it should be that $\| A\mathbf{u} \| = \| \mathbf{u} \|$ and $\| A\mathbf{v} \| = \| \mathbf{v} \|$.

c. Answers will again vary, but the cosines should be equal.

d. A conjecture is that multiplying by A does not change the lengths of vectors or the angles between vectors.

33. **[M]** Answers to the calculations will vary, but will demonstrate that the mapping $\mathbf{x} \mapsto T(\mathbf{x}) = \left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}$ (for $\mathbf{v} \neq \mathbf{0}$) is a linear transformation. To confirm this, let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$, and let c be any scalar. Then

$$T(\mathbf{x}+\mathbf{y}) = \left(\frac{(\mathbf{x}+\mathbf{y})\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} = \left(\frac{(\mathbf{x}\cdot\mathbf{v})+(\mathbf{y}\cdot\mathbf{v})}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} = \left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} + \left(\frac{\mathbf{y}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} = T(\mathbf{x}) + T(\mathbf{y})$$

and

$$T(c\mathbf{x}) = \left(\frac{(c\mathbf{x})\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} = \left(\frac{c(\mathbf{x}\cdot\mathbf{v})}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} = c\left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} = cT(\mathbf{x})$$

34. **[M]** One finds that $N = \begin{bmatrix} -5 & 1 \\ -1 & 4 \\ 1 & 0 \\ 0 & -1 \\ 0 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 5 & 0 & -1/3 \\ 0 & 1 & 1 & 0 & -4/3 \\ 0 & 0 & 0 & 1 & 1/3 \end{bmatrix}$.

The row-column rule for computing RN produces the 3×2 zero matrix, which shows that the rows of R are orthogonal to the columns of N. This is expected by Theorem 3 since each row of R is in Row A and each column of N is in Nul A.

6.2 SOLUTIONS

Notes: The nonsquare matrices in Theorems 6 and 7 are needed for the QR factorization in Section 6.4. It is important to emphasize that the term *orthogonal matrix* applies only to certain *square* matrices. The subsection on orthogonal projections not only sets the stage for the general case in Section 6.3, it also provides what is needed for the orthogonal diagonalization exercises in Section 7.1, because none of the eigenspaces there have dimension greater than 2. For this reason, the Gram-Schmidt process (Section 6.4) is not really needed in Chapter 7. Exercises 13 and 14 are good preparation for Section 6.3.

1. Since $\begin{bmatrix} -1 \\ 4 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ -4 \\ -7 \end{bmatrix} = 2 \neq 0$, the set is not orthogonal.

2. Since $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix} = 0$, the set is orthogonal.

3. Since $\begin{bmatrix} -6 \\ -3 \\ 9 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} = -30 \neq 0$, the set is not orthogonal.

4. Since $\begin{bmatrix} 2 \\ -5 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -5 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix} = 0,$ the set is orthogonal.

5. Since $\begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix} = 0,$ the set is orthogonal.

6. Since $\begin{bmatrix} -4 \\ 1 \\ -3 \\ 8 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 3 \\ 5 \\ -1 \end{bmatrix} = -32 \neq 0,$ the set is not orthogonal.

7. Since $\mathbf{u}_1 \cdot \mathbf{u}_2 = 12 - 12 = 0$, $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal set. Since the vectors are non-zero, $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent by Theorem 4. Two such vectors in $\mathbb{R}^2$ automatically form a basis for $\mathbb{R}^2$. So $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $\mathbb{R}^2$. By Theorem 5,

$$\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{x} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = 3\mathbf{u}_1 + \frac{1}{2}\mathbf{u}_2$$

8. Since $\mathbf{u}_1 \cdot \mathbf{u}_2 = -6 + 6 = 0$, $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal set. Since the vectors are non-zero, $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent by Theorem 4. Two such vectors in $\mathbb{R}^2$ automatically form a basis for $\mathbb{R}^2$. So $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $\mathbb{R}^2$. By Theorem 5,

$$\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{x} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = -\frac{3}{2}\mathbf{u}_1 + \frac{3}{4}\mathbf{u}_2$$

9. Since $\mathbf{u}_1 \cdot \mathbf{u}_2 = \mathbf{u}_1 \cdot \mathbf{u}_3 = \mathbf{u}_2 \cdot \mathbf{u}_3 = 0$, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set. Since the vectors are non-zero, $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ are linearly independent by Theorem 4. Three such vectors in $\mathbb{R}^3$ automatically form a basis for $\mathbb{R}^3$. So $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^3$. By Theorem 5,

$$\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{x} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{x} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3}\mathbf{u}_3 = \frac{5}{2}\mathbf{u}_1 - \frac{3}{2}\mathbf{u}_2 + 2\mathbf{u}_3$$

10. Since $\mathbf{u}_1 \cdot \mathbf{u}_2 = \mathbf{u}_1 \cdot \mathbf{u}_3 = \mathbf{u}_2 \cdot \mathbf{u}_3 = 0$, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set. Since the vectors are non-zero, $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ are linearly independent by Theorem 4. Three such vectors in $\mathbb{R}^3$ automatically form a basis for $\mathbb{R}^3$. So $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^3$. By Theorem 5,

$$\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{x} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{x} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3}\mathbf{u}_3 = \frac{4}{3}\mathbf{u}_1 + \frac{1}{3}\mathbf{u}_2 + \frac{1}{3}\mathbf{u}_3$$

11. Let $\mathbf{y}=\begin{bmatrix}1\\7\end{bmatrix}$ and $\mathbf{u}=\begin{bmatrix}-4\\2\end{bmatrix}$. The orthogonal projection of **y** onto the line through **u** and the origin is the orthogonal projection of **y** onto **u**, and this vector is $\hat{\mathbf{y}}=\dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}=\dfrac{1}{2}\mathbf{u}=\begin{bmatrix}-2\\1\end{bmatrix}$.

12. Let $\mathbf{y}=\begin{bmatrix}1\\-1\end{bmatrix}$ and $\mathbf{u}=\begin{bmatrix}-1\\3\end{bmatrix}$. The orthogonal projection of **y** onto the line through **u** and the origin is the orthogonal projection of **y** onto **u**, and this vector is $\hat{\mathbf{y}}=\dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}=-\dfrac{2}{5}\mathbf{u}=\begin{bmatrix}2/5\\-6/5\end{bmatrix}$.

13. The orthogonal projection of **y** onto **u** is $\hat{\mathbf{y}}=\dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}=-\dfrac{13}{65}\mathbf{u}=\begin{bmatrix}-4/5\\7/5\end{bmatrix}$. The component of **y** orthogonal to **u** is $\mathbf{y}-\hat{\mathbf{y}}=\begin{bmatrix}14/5\\8/5\end{bmatrix}$. Thus $\mathbf{y}=\hat{\mathbf{y}}+(\mathbf{y}-\hat{\mathbf{y}})=\begin{bmatrix}-4/5\\7/5\end{bmatrix}+\begin{bmatrix}14/5\\8/5\end{bmatrix}$.

14. The orthogonal projection of **y** onto **u** is $\hat{\mathbf{y}}=\dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}=\dfrac{2}{5}\mathbf{u}=\begin{bmatrix}14/5\\2/5\end{bmatrix}$. The component of **y** orthogonal to **u** is $\mathbf{y}-\hat{\mathbf{y}}=\begin{bmatrix}-4/5\\28/5\end{bmatrix}$. Thus $\mathbf{y}=\hat{\mathbf{y}}+(\mathbf{y}-\hat{\mathbf{y}})=\begin{bmatrix}14/5\\2/5\end{bmatrix}+\begin{bmatrix}-4/5\\28/5\end{bmatrix}$.

15. The distance from **y** to the line through **u** and the origin is $\|\mathbf{y}-\hat{\mathbf{y}}\|$. One computes that

$\mathbf{y}-\hat{\mathbf{y}}=\mathbf{y}-\dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}=\begin{bmatrix}3\\1\end{bmatrix}-\dfrac{3}{10}\begin{bmatrix}8\\6\end{bmatrix}=\begin{bmatrix}3/5\\-4/5\end{bmatrix}$, so $\|\mathbf{y}-\hat{\mathbf{y}}\|=\sqrt{9/25+16/25}=1$ is the desired distance.

16. The distance from **y** to the line through **u** and the origin is $\|\mathbf{y}-\hat{\mathbf{y}}\|$. One computes that

$\mathbf{y}-\hat{\mathbf{y}}=\mathbf{y}-\dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}=\begin{bmatrix}-3\\9\end{bmatrix}-3\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}-6\\3\end{bmatrix}$, so $\|\mathbf{y}-\hat{\mathbf{y}}\|=\sqrt{36+9}=3\sqrt{5}$ is the desired distance.

17. Let $\mathbf{u}=\begin{bmatrix}1/3\\1/3\\1/3\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}-1/2\\0\\1/2\end{bmatrix}$. Since $\mathbf{u}\cdot\mathbf{v}=0$, $\{\mathbf{u},\mathbf{v}\}$ is an orthogonal set. However, $\|\mathbf{u}\|^2=\mathbf{u}\cdot\mathbf{u}=1/3$ and $\|\mathbf{v}\|^2=\mathbf{v}\cdot\mathbf{v}=1/2$, so $\{\mathbf{u},\mathbf{v}\}$ is not an orthonormal set. The vectors **u** and **v** may be normalized to form the orthonormal set

$$\left\{\frac{\mathbf{u}}{\|\mathbf{u}\|},\frac{\mathbf{v}}{\|\mathbf{v}\|}\right\}=\left\{\begin{bmatrix}\sqrt{3}/3\\\sqrt{3}/3\\\sqrt{3}/3\end{bmatrix},\begin{bmatrix}-\sqrt{2}/2\\0\\\sqrt{2}/2\end{bmatrix}\right\}$$

18. Let $\mathbf{u}=\begin{bmatrix}0\\1\\0\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}0\\-1\\0\end{bmatrix}$. Since $\mathbf{u}\cdot\mathbf{v}=-1\neq 0$, $\{\mathbf{u},\mathbf{v}\}$ is not an orthogonal set.

19. Let $\mathbf{u}=\begin{bmatrix}-.6\\.8\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}.8\\.6\end{bmatrix}$. Since **u** · **v** = 0, {**u**, **v**} is an orthogonal set. Also, $\|\mathbf{u}\|^2=\mathbf{u}\cdot\mathbf{u}=1$ and $\|\mathbf{v}\|^2=\mathbf{v}\cdot\mathbf{v}=1$, so {**u**, **v**} is an orthonormal set.

20. Let $\mathbf{u}=\begin{bmatrix}-2/3\\1/3\\2/3\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}1/3\\2/3\\0\end{bmatrix}$. Since **u** · **v** = 0, {**u**, **v**} is an orthogonal set. However, $\|\mathbf{u}\|^2=\mathbf{u}\cdot\mathbf{u}=1$ and $\|\mathbf{v}\|^2=\mathbf{v}\cdot\mathbf{v}=5/9$, so {**u**, **v**} is not an orthonormal set. The vectors **u** and **v** may be normalized to form the orthonormal set $\left\{\frac{\mathbf{u}}{\|\mathbf{u}\|},\frac{\mathbf{v}}{\|\mathbf{v}\|}\right\}=\left\{\begin{bmatrix}-2/3\\1/3\\2/3\end{bmatrix},\begin{bmatrix}1/\sqrt{5}\\2/\sqrt{5}\\0\end{bmatrix}\right\}$.

21. Let $\mathbf{u}=\begin{bmatrix}1/\sqrt{10}\\3/\sqrt{20}\\3/\sqrt{20}\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}3/\sqrt{10}\\-1/\sqrt{20}\\-1/\sqrt{20}\end{bmatrix}$, and $\mathbf{w}=\begin{bmatrix}0\\-1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}$. Since **u** · **v** = **u** · **w** = **v** · **w** = 0, {**u**, **v**, **w**} is an orthogonal set. Also, $\|\mathbf{u}\|^2=\mathbf{u}\cdot\mathbf{u}=1$, $\|\mathbf{v}\|^2=\mathbf{v}\cdot\mathbf{v}=1$, and $\|\mathbf{w}\|^2=\mathbf{w}\cdot\mathbf{w}=1$, so {**u**, **v**, **w**} is an orthonormal set.

22. Let $\mathbf{u}=\begin{bmatrix}1/\sqrt{18}\\4/\sqrt{18}\\1/\sqrt{18}\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\end{bmatrix}$, and $\mathbf{w}=\begin{bmatrix}-2/3\\1/3\\-2/3\end{bmatrix}$. Since **u** · **v** = **u** · **w** = **v** · **w** = 0, {**u**, **v**, **w**} is an orthogonal set. Also, $\|\mathbf{u}\|^2=\mathbf{u}\cdot\mathbf{u}=1$, $\|\mathbf{v}\|^2=\mathbf{v}\cdot\mathbf{v}=1$, and $\|\mathbf{w}\|^2=\mathbf{w}\cdot\mathbf{w}=1$, so {**u**, **v**, **w**} is an orthonormal set.

23. **a**. True. For example, the vectors **u** and **y** in Example 3 are linearly independent but not orthogonal.
b. True. The formulas for the weights are given in Theorem 5.
c. False. See the paragraph following Example 5.
d. False. The matrix must also be square. See the paragraph before Example 7.
e. False. See Example 4. The distance is $\|\mathbf{y}-\hat{\mathbf{y}}\|$.

24. **a**. True. But every orthogonal set of *nonzero vectors* is linearly independent. See Theorem 4.
b. False. To be orthonormal, the vectors is *S* must be unit vectors as well as being orthogonal to each other.
c. True. See Theorem 7(a).
d. True. See the paragraph before Example 3.
e. True. See the paragraph before Example 7.

25. To prove part (b), note that

$$(U\mathbf{x})\cdot(U\mathbf{y})=(U\mathbf{x})^T(U\mathbf{y})=\mathbf{x}^TU^TU\mathbf{y}=\mathbf{x}^T\mathbf{y}=\mathbf{x}\cdot\mathbf{y}$$

because $U^T U = I$. If $\mathbf{y} = \mathbf{x}$ in part (b), $(U\mathbf{x}) \cdot (U\mathbf{x}) = \mathbf{x} \cdot \mathbf{x}$, which implies part (a). Part (c) of the Theorem follows immediately fom part (b).

26. A set of n nonzero orthogonal vectors must be linearly independent by Theorem 4, so if such a set spans W it is a basis for W. Thus W is an n-dimensional subspace of $\mathbb{R}^n$, and $W = \mathbb{R}^n$.

27. If U has orthonormal columns, then $U^T U = I$ by Theorem 6. If U is also a square matrix, then the equation $U^T U = I$ implies that U is invertible by the Invertible Matrix Theorem.

28. If U is an $n \times n$ orthogonal matrix, then $I = UU^{-1} = UU^T$. Since U is the transpose of U^T, Theorem 6 applied to U^T says that U^T has ~~orthogonal~~ columns. In particular, the columns of U^T are linearly independent and hence form a basis for $\mathbb{R}^n$ by the Invertible Matrix Theorem. That is, the rows of U form a basis (an orthonormal basis) for $\mathbb{R}^n$.

29. Since U and V are orthogonal, each is invertible. By Theorem 6 in Section 2.2, UV is invertible and $(UV)^{-1} = V^{-1}U^{-1} = V^T U^T = (UV)^T$, where the final equality holds by Theorem 3 in Section 2.1. Thus UV is an orthogonal matrix.

30. If U is an orthogonal matrix, its columns are orthonormal. Interchanging the columns does not change their orthonormality, so the new matrix – say, V – still has orthonormal columns. By Theorem 6, $V^T V = I$. Since V is square, $V^T = V^{-1}$ by the Invertible Matrix Theorem.

31. Suppose that $\hat{\mathbf{y}} = \dfrac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$. Replacing $\mathbf{u}$ by $c\mathbf{u}$ with $c \neq 0$ gives

$$\frac{\mathbf{y} \cdot (c\mathbf{u})}{(c\mathbf{u}) \cdot (c\mathbf{u})}(c\mathbf{u}) = \frac{c(\mathbf{y} \cdot \mathbf{u})}{c^2(\mathbf{u} \cdot \mathbf{u})}(c)\mathbf{u} = \frac{c^2(\mathbf{y} \cdot \mathbf{u})}{c^2(\mathbf{u} \cdot \mathbf{u})}\mathbf{u} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} = \hat{\mathbf{y}}$$

So $\hat{\mathbf{y}}$ does not depend on the choice of a nonzero $\mathbf{u}$ in the line L used in the formula.

32. If $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, then by Theorem 1(c) in Section 6.1,

$$(c_1\mathbf{v}_1) \cdot (c_2\mathbf{v}_2) = c_1[\mathbf{v}_1 \cdot (c_2\mathbf{v}_2)] = c_1c_2(\mathbf{v}_1 \cdot \mathbf{v}_2) = c_1c_2 0 = 0$$

33. Let $L = \text{Span}\{\mathbf{u}\}$, where $\mathbf{u}$ is nonzero, and let $T(\mathbf{x}) = \dfrac{\mathbf{x} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$. For any vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and any scalars c and d, the properties of the inner product (Theorem 1) show that

$$\begin{aligned} T(c\mathbf{x} + d\mathbf{y}) &= \frac{(c\mathbf{x} + d\mathbf{y}) \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} \\ &= \frac{c\mathbf{x} \cdot \mathbf{u} + d\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} \\ &= \frac{c\mathbf{x} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} + \frac{d\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} \\ &= cT(\mathbf{x}) + dT(\mathbf{y}) \end{aligned}$$

Thus T is a linear transformation. Another approach is to view T as the composition of the following three linear mappings: $\mathbf{x} \mapsto a = \mathbf{x} \cdot \mathbf{v}$, $a \mapsto b = a / \mathbf{v} \cdot \mathbf{v}$, and $b \mapsto b\mathbf{v}$.

34. Let $L = \text{Span}\{\mathbf{u}\}$, where **u** is nonzero, and let $T(\mathbf{x}) = \text{refl}_L\mathbf{y} = 2\text{proj}_L\mathbf{y} - \mathbf{y}$. By Exercise 33, the mapping $\mathbf{y} \mapsto \text{proj}_L\mathbf{y}$ is linear. Thus for any vectors **y** and **z** in $\mathbb{R}^n$ and any scalars *c* and *d*,

$$\begin{aligned} T(c\mathbf{y} + d\mathbf{z}) &= 2\ \text{proj}_L(c\mathbf{y} + d\mathbf{z}) - (c\mathbf{y} + d\mathbf{z}) \\ &= 2(c\ \text{proj}_L\mathbf{y} + d\ \text{proj}_L\mathbf{z}) - c\mathbf{y} - d\mathbf{z} \\ &= 2c\ \text{proj}_L\mathbf{y} - c\mathbf{y} + 2d\ \text{proj}_L\mathbf{z} - d\mathbf{z} \\ &= c(2\ \text{proj}_L\mathbf{y} - \mathbf{y}) + d(2\ \text{proj}_L\mathbf{z} - \mathbf{z}) \\ &= cT(\mathbf{y}) + dT(\mathbf{z}) \end{aligned}$$

Thus *T* is a linear transformation.

35. **[M]** One can compute that $A^TA = 100I_4$. Since the off-diagonal entries in A^TA are zero, the columns of *A* are orthogonal.

36. **[M]**

a. One computes that $U^TU = I_4$, while

$$UU^T = \left(\frac{1}{100}\right)\begin{bmatrix} 82 & 0 & -20 & 8 & 6 & 20 & 24 & 0 \\ 0 & 42 & 24 & 0 & -20 & 6 & 20 & -32 \\ -20 & 24 & 58 & 20 & 0 & 32 & 0 & 6 \\ 8 & 0 & 20 & 82 & 24 & -20 & 6 & 0 \\ 6 & -20 & 0 & 24 & 18 & 0 & -8 & 20 \\ 20 & 6 & 32 & -20 & 0 & 58 & 0 & 24 \\ 24 & 20 & 0 & 6 & -8 & 0 & 18 & -20 \\ 0 & -32 & 6 & 0 & 20 & 24 & -20 & 42 \end{bmatrix}$$

The matrices U^TU and UU^T are of different sizes and look nothing like each other.

b. Answers will vary. The vector $\mathbf{p} = UU^T\mathbf{y}$ is in Col *U* because $\mathbf{p} = U(U^T\mathbf{y})$. Since the columns of *U* are simply scaled versions of the columns of *A*, Col *U* = Col *A*. Thus each **p** is in Col *A*.

c. One computes that $U^T\mathbf{z} = \mathbf{0}$.

d. From (c), **z** is orthogonal to each column of *A*. By Exercise 29 in Section 6.1, **z** must be orthogonal to every vector in Col *A*; that is, **z** is in $(\text{Col}\ A)^{\perp}$.

6.3 SOLUTIONS

Notes: Example 1 seems to help students understand Theorem 8. Theorem 8 is needed for the Gram-Schmidt process (but only for a subspace that itself has an orthogonal basis). Theorems 8 and 9 are needed for the discussions of least squares in Sections 6.5 and 6.6. Theorem 10 is used with the QR factorization to provide a good numerical method for solving least squares problems, in Section 6.5. Exercises 19 and 20 lead naturally into consideration of the Gram-Schmidt process.

1. The vector in $\text{Span}\{\mathbf{u}_4\}$ is $\dfrac{\mathbf{x}\cdot\mathbf{u}_4}{\mathbf{u}_4\cdot\mathbf{u}_4}\mathbf{u}_4 = \dfrac{72}{36}\mathbf{u}_4 = 2\mathbf{u}_4 = \begin{bmatrix}10\\-6\\-2\\2\end{bmatrix}$. Since

$\mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + \dfrac{\mathbf{x}\cdot\mathbf{u}_4}{\mathbf{u}_4\cdot\mathbf{u}_4}\mathbf{u}_4$, the vector $\mathbf{x} - \dfrac{\mathbf{x}\cdot\mathbf{u}_4}{\mathbf{u}_4\cdot\mathbf{u}_4}\mathbf{u}_4 = \begin{bmatrix}10\\-8\\2\\0\end{bmatrix} - \begin{bmatrix}10\\-6\\-2\\2\end{bmatrix} = \begin{bmatrix}0\\-2\\4\\-2\end{bmatrix}$ is in

$\text{Span}\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$.

2. The vector in $\text{Span}\{\mathbf{u}_1\}$ is $\dfrac{\mathbf{v}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 = \dfrac{14}{7}\mathbf{u}_1 = 2\mathbf{u}_1 = \begin{bmatrix}2\\4\\2\\2\end{bmatrix}$. Since $\mathbf{x} = \dfrac{\mathbf{v}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + c_4\mathbf{u}_4$,

the vector $\mathbf{v} - \dfrac{\mathbf{v}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 = \begin{bmatrix}4\\5\\-3\\3\end{bmatrix} - \begin{bmatrix}2\\4\\2\\2\end{bmatrix} = \begin{bmatrix}2\\1\\-5\\1\end{bmatrix}$ is in $\text{Span}\{\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4\}$.

3. Since $\mathbf{u}_1\cdot\mathbf{u}_2 = -1+1+0 = 0$, $\{\mathbf{u}_1,\mathbf{u}_2\}$ is an orthogonal set. The orthogonal projection of **y** onto

$\text{Span}\{\mathbf{u}_1,\mathbf{u}_2\}$ is $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \dfrac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \dfrac{3}{2}\mathbf{u}_1 + \dfrac{5}{2}\mathbf{u}_2 = \dfrac{3}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} + \dfrac{5}{2}\begin{bmatrix}-1\\1\\0\end{bmatrix} = \begin{bmatrix}-1\\4\\0\end{bmatrix}$.

4. Since $\mathbf{u}_1\cdot\mathbf{u}_2 = -12+12+0 = 0$, $\{\mathbf{u}_1,\mathbf{u}_2\}$ is an orthogonal set. The orthogonal projection of **y** onto

$\text{Span}\{\mathbf{u}_1,\mathbf{u}_2\}$ is $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \dfrac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \dfrac{30}{25}\mathbf{u}_1 - \dfrac{15}{25}\mathbf{u}_2 = \dfrac{6}{5}\begin{bmatrix}3\\4\\0\end{bmatrix} - \dfrac{3}{5}\begin{bmatrix}-4\\3\\0\end{bmatrix} = \begin{bmatrix}6\\3\\0\end{bmatrix}$.

5. Since $\mathbf{u}_1\cdot\mathbf{u}_2 = 3+1-4=0$, $\{\mathbf{u}_1,\mathbf{u}_2\}$ is an orthogonal set. The orthogonal projection of **y** onto Span$\{\mathbf{u}_1,\mathbf{u}_2\}$ is $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \dfrac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \dfrac{7}{14}\mathbf{u}_1 - \dfrac{15}{6}\mathbf{u}_2 = \dfrac{1}{2}\begin{bmatrix}3\\-1\\2\end{bmatrix} - \dfrac{5}{2}\begin{bmatrix}1\\-1\\-2\end{bmatrix} = \begin{bmatrix}-1\\2\\6\end{bmatrix}$.

6. Since $\mathbf{u}_1\cdot\mathbf{u}_2 = 0-1+1=0$, $\{\mathbf{u}_1,\mathbf{u}_2\}$ is an orthogonal set. The orthogonal projection of **y** onto Span$\{\mathbf{u}_1,\mathbf{u}_2\}$ is $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \dfrac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = -\dfrac{27}{18}\mathbf{u}_1 + \dfrac{5}{2}\mathbf{u}_2 = -\dfrac{3}{2}\begin{bmatrix}-4\\-1\\1\end{bmatrix} + \dfrac{5}{2}\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}6\\4\\1\end{bmatrix}$.

7. Since $\mathbf{u}_1\cdot\mathbf{u}_2 = 5+3-8=0$, $\{\mathbf{u}_1,\mathbf{u}_2\}$ is an orthogonal set. By the Orthogonal Decomposition Theorem, $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \dfrac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = 0\mathbf{u}_1 + \dfrac{2}{3}\mathbf{u}_2 = \begin{bmatrix}10/3\\2/3\\8/3\end{bmatrix}, \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix}-7/3\\7/3\\7/3\end{bmatrix}$ and $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, where $\hat{\mathbf{y}}$ is in W and **z** is in $W^\perp$.

8. Since $\mathbf{u}_1\cdot\mathbf{u}_2 = -1+3-2=0$, $\{\mathbf{u}_1,\mathbf{u}_2\}$ is an orthogonal set. By the Orthogonal Decomposition Theorem, $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \dfrac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = 2\mathbf{u}_1 + \dfrac{1}{2}\mathbf{u}_2 = \begin{bmatrix}3/2\\7/2\\1\end{bmatrix}, \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix}-5/2\\1/2\\2\end{bmatrix}$ and $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, where $\hat{\mathbf{y}}$ is in W and **z** is in $W^\perp$.

9. Since $\mathbf{u}_1\cdot\mathbf{u}_2 = \mathbf{u}_1\cdot\mathbf{u}_3 = \mathbf{u}_2\cdot\mathbf{u}_3 = 0$, $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ is an orthogonal set. By the Orthogonal Decomposition Theorem,

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{y}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3}\mathbf{u}_3 = 2\mathbf{u}_1 + \frac{2}{3}\mathbf{u}_2 - \frac{2}{3}\mathbf{u}_3 = \begin{bmatrix}2\\4\\0\\0\end{bmatrix}, \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix}2\\-1\\3\\-1\end{bmatrix}$$ and $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, where $\hat{\mathbf{y}}$ is in W and **z** is in $W^\perp$.

10. Since $\mathbf{u}_1\cdot\mathbf{u}_2 = \mathbf{u}_1\cdot\mathbf{u}_3 = \mathbf{u}_2\cdot\mathbf{u}_3 = 0$, $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ is an orthogonal set. By the Orthogonal Decomposition Theorem,

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{y}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3}\mathbf{u}_3 = \frac{1}{3}\mathbf{u}_1 + \frac{14}{3}\mathbf{u}_2 - \frac{5}{3}\mathbf{u}_3 = \begin{bmatrix}5\\2\\3\\6\end{bmatrix}, \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix}-2\\2\\2\\0\end{bmatrix}$$ and $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, where $\hat{\mathbf{y}}$ is in W and **z** is in $W^\perp$.

11. Note that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. The Best Approximation Theorem says that $\hat{\mathbf{y}}$, which is the orthogonal projection of **y** onto $W = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, is the closest point to **y** in *W*. This vector is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{y}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 = \frac{1}{2}\mathbf{v}_1 + \frac{3}{2}\mathbf{v}_2 = \begin{bmatrix} 3 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$

12. Note that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. The Best Approximation Theorem says that $\hat{\mathbf{y}}$, which is the orthogonal projection of *y* onto $W = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, is the closest point to **y** in *W*. This vector is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{y}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 = 3\mathbf{v}_1 + 1\mathbf{v}_2 = \begin{bmatrix} -1 \\ -5 \\ -3 \\ 9 \end{bmatrix}.$$

13. Note that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. By the Best Approximation Theorem, the closest point in

$\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ to **z** is $\hat{\mathbf{z}} = \dfrac{\mathbf{z}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \dfrac{\mathbf{z}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 = \dfrac{2}{3}\mathbf{v}_1 - \dfrac{7}{3}\mathbf{v}_2 = \begin{bmatrix} -1 \\ -3 \\ -2 \\ 3 \end{bmatrix}.$

14. Note that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. By the Best Approximation Theorem, the closest point in

$\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ to **z** is $\hat{\mathbf{z}} = \dfrac{\mathbf{z}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \dfrac{\mathbf{z}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 = \dfrac{1}{2}\mathbf{v}_1 + 0\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ -1/2 \\ -3/2 \end{bmatrix}.$

15. The distance from the point **y** in $\mathbb{R}^3$ to a subspace *W* is defined as the distance from **y** to the closest point in *W*. Since the closest point in *W* to **y** is $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$, the desired distance is $\|\mathbf{y} - \hat{\mathbf{y}}\|$. One computes that $\hat{\mathbf{y}} = \begin{bmatrix} 3 \\ -9 \\ -1 \end{bmatrix}, \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 2 \\ 0 \\ 6 \end{bmatrix}$, and $\|\mathbf{y} - \hat{\mathbf{y}}\| = \sqrt{40} = 2\sqrt{10}$.

16. The distance from the point **y** in $\mathbb{R}^4$ to a subspace *W* is defined as the distance from **y** to the closest point in *W*. Since the closest point in *W* to **y** is $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$, the desired distance is $\|\mathbf{y} - \hat{\mathbf{y}}\|$. One computes that $\hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ -3 \\ 9 \end{bmatrix}, \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix}$, and $\|\mathbf{y} - \hat{\mathbf{y}}\| = 8$.

17. **a**. $U^TU = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $UU^T = \begin{bmatrix} 8/9 & -2/9 & 2/9 \\ -2/9 & 5/9 & 4/9 \\ 2/9 & 4/9 & 5/9 \end{bmatrix}$.

b. Since $U^TU = I_2$, the columns of U form an orthonormal basis for W, and by Theorem 10

$$\text{proj}_W \mathbf{y} = UU^T\mathbf{y} = \begin{bmatrix} 8/9 & -2/9 & 2/9 \\ -2/9 & 5/9 & 4/9 \\ 2/9 & 4/9 & 5/9 \end{bmatrix}\begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}.$$

18. **a**. $U^TU = [1] = 1$, $UU^T = \begin{bmatrix} 1/10 & -3/10 \\ -3/10 & 9/10 \end{bmatrix}$

b. Since $U^TU = 1$, $\{\mathbf{u}_1\}$ forms an orthonormal basis for W, and by Theorem 10

$$\text{proj}_W \mathbf{y} = UU^T\mathbf{y} = \begin{bmatrix} 1/10 & -3/10 \\ -3/10 & 9/10 \end{bmatrix}\begin{bmatrix} 7 \\ 9 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}.$$

19. By the Orthogonal Decomposition Theorem, $\mathbf{u}_3$ is the sum of a vector in $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ and a vector $\mathbf{v}$ orthogonal to W. This exercise asks for the vector $\mathbf{v}$:

$\mathbf{v} = \mathbf{u}_3 - \text{proj}_W \mathbf{u}_3 = \mathbf{u}_3 - \left(-\frac{1}{3}\mathbf{u}_1 + \frac{1}{15}\mathbf{u}_2\right) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ -2/5 \\ 4/5 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/5 \\ 1/5 \end{bmatrix}$. Any multiple of the vector $\mathbf{v}$ will also be in $W^\perp$.

20. By the Orthogonal Decomposition Theorem, $\mathbf{u}_4$ is the sum of a vector in $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ and a vector $\mathbf{v}$ orthogonal to W. This exercise asks for the vector $\mathbf{v}$:

$\mathbf{v} = \mathbf{u}_4 - \text{proj}_W \mathbf{u}_4 = \mathbf{u}_4 - \left(\frac{1}{6}\mathbf{u}_1 - \frac{1}{30}\mathbf{u}_2\right) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1/5 \\ -2/5 \end{bmatrix} = \begin{bmatrix} 0 \\ 4/5 \\ 2/5 \end{bmatrix}$. Any multiple of the vector $\mathbf{v}$ will also be in $W^\perp$.

21. **a**. True. See the calculations for $\mathbf{z}_2$ in Example 1 or the box after Example 6 in Section 6.1.

b. True. See the Orthogonal Decomposition Theorem.

c. False. See the last paragraph in the proof of Theorem 8, or see the second paragraph after the statement of Theorem 9.

d. True. See the box before the Best Approximation Theorem.

e. True. Theorem 10 applies to the column space W of U because the columns of U are linearly independent and hence form a basis for W.

22. **a**. True. See the proof of the Orthogonal Decomposition Theorem.

b. True. See the subsection "A Geometric Interpretation of the Orthogonal Projection."

c. True. The orthgonal decomposition in Theorem 8 is unique.

d. False. The Best Approximation Theorem says that the best approximation to $\mathbf{y}$ is $\text{proj}_W \mathbf{y}$.

e. False. This statement is only true if **x** is in the column space of U. If $n > p$, then the column space of U will not be all of $\mathbb{R}^n$, so the statement cannot be true for all **x** in $\mathbb{R}^n$.

23. By the Orthogonal Decomposition Theorem, each **x** in $\mathbb{R}^n$ can be written uniquely as $\mathbf{x} = \mathbf{p} + \mathbf{u}$, with **p** in Row A and **u** in $(\text{Row}\,A)^{\perp}$. By Theorem 3 in Section 6.1, $(\text{Row}\,A)^{\perp} = \text{Nul}\,A$, so **u** is in Nul$A$.

Next, suppose $A\mathbf{x} = \mathbf{b}$ is consistent. Let **x** be a solution and write $\mathbf{x} = \mathbf{p} + \mathbf{u}$ as above. Then $A\mathbf{p} = A(\mathbf{x} - \mathbf{u}) = A\mathbf{x} - A\mathbf{u} = \mathbf{b} - \mathbf{0} = \mathbf{b}$, so the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution **p** in Row A. Finally, suppose that **p** and $\mathbf{p}_1$ are both in Row$A$ and both satisfy $A\mathbf{x} = \mathbf{b}$. Then $\mathbf{p} - \mathbf{p}_1$ is in $\text{Nul}\,A = (\text{Row}\,A)^{\perp}$, since $A(\mathbf{p} - \mathbf{p}_1) = A\mathbf{p} - A\mathbf{p}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0}$. The equations $\mathbf{p} = \mathbf{p}_1 + (\mathbf{p} - \mathbf{p}_1)$ and $\mathbf{p} = \mathbf{p} + \mathbf{0}$ both then decompose **p** as the sum of a vector in Row$A$ and a vector in $(\text{Row}\,A)^{\perp}$. By the uniqueness of the orthogonal decomposition (Theorem 8), $\mathbf{p} = \mathbf{p}_1$, and **p** is unique.

24. **a**. By hypothesis, the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_p$ are pairwise orthogonal, and the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_q$ are pairwise orthogonal. Since $\mathbf{w}_i$ is in W for any i and $\mathbf{v}_j$ is in $W^{\perp}$ for any j, $\mathbf{w}_i \cdot \mathbf{v}_j = 0$ for any i and j. Thus $\{\mathbf{w}_1,\ldots,\mathbf{w}_p,\mathbf{v}_1,\ldots,\mathbf{v}_q\}$ forms an orthogonal set.

b. For any **y** in $\mathbb{R}^n$, write $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ as in the Orthogonal Decomposition Theorem, with $\hat{\mathbf{y}}$ in W and **z** in $W^{\perp}$. Then there exist scalars $c_1,\ldots,c_p$ and $d_1,\ldots,d_q$ such that $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} = c_1\mathbf{w}_1 + \ldots + c_p\mathbf{w}_p + d_1\mathbf{v}_1 + \ldots + d_q\mathbf{v}_q$. Thus the set $\{\mathbf{w}_1,\ldots,\mathbf{w}_p,\mathbf{v}_1,\ldots,\mathbf{v}_q\}$ spans $\mathbb{R}^n$.

c. The set $\{\mathbf{w}_1,\ldots,\mathbf{w}_p,\mathbf{v}_1,\ldots,\mathbf{v}_q\}$ is linearly independent by (a) and spans $\mathbb{R}^n$ by (b), and is thus a basis for $\mathbb{R}^n$. Hence $\dim W + \dim W^{\perp} = p + q = \dim \mathbb{R}^n$.

25. **[M]** Since $U^TU = I_4$, U has orthonormal columns by Theorem 6 in Section 6.2. The closest point to **y** in Col U is the orthogonal projection $\hat{\mathbf{y}}$ of **y** onto Col U. From Theorem 10, $\hat{\mathbf{y}} = UU^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} 1.2 \\ .4 \\ 1.2 \\ 1.2 \\ .4 \\ 1.2 \\ .4 \\ .4 \end{bmatrix}$

26.[M] The distance from $\mathbf{b}$ to $\operatorname{Col} U$ is $\|\mathbf{b}-\hat{\mathbf{b}}\|$, where $\hat{\mathbf{b}} = UU^{\mathrm{T}}\mathbf{b}$. One computes that

$$\hat{\mathbf{b}} = UU^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} .2 \\ .92 \\ .44 \\ 1 \\ -.2 \\ -.44 \\ .6 \\ -.92 \end{bmatrix}, \mathbf{b}-\hat{\mathbf{b}} = \begin{bmatrix} .8 \\ .08 \\ .56 \\ 0 \\ -.8 \\ -.56 \\ -1.6 \\ -.08 \end{bmatrix}, \|\mathbf{b}-\hat{\mathbf{b}}\| = \frac{\sqrt{112}}{5}.$$

which is 2.1166 to four decimal places.

6.4 SOLUTIONS

Notes: The QR factorization encapsulates the essential outcome of the Gram-Schmidt process, just as the LU factorization describes the result of a row reduction process. For practical use of linear algebra, the factorizations are more important than the algorithms that produce them. In fact, the Gram-Schmidt process is *not* the appropriate way to compute the QR factorization. For that reason, one should consider deemphasizing the hand calculation of the Gram-Schmidt process, even though it provides easy exam questions.

The Gram-Schmidt process is used in Sections 6.7 and 6.8, in connection with various sets of orthogonal polynomials. The process is mentioned in Sections 7.1 and 7.4, but the one-dimensional projection constructed in Section 6.2 will suffice. The QR factorization is used in an optional subsection of Section 6.5, and it is needed in Supplementary Exercise 7 of Chapter 7 to produce the Cholesky factorization of a positive definite matrix.

1. Set $\mathbf{v}_1 = \mathbf{x}_1$ and compute that $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - 3\mathbf{v}_1 = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}$. Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix} \right\}$.

2. Set $\mathbf{v}_1 = \mathbf{x}_1$ and compute that $\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - \frac{1}{2}\mathbf{v}_1 = \begin{bmatrix} 5 \\ 4 \\ -8 \end{bmatrix}$. Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 4 \\ -8 \end{bmatrix} \right\}$.

3. Set $\mathbf{v}_1 = \mathbf{x}_1$ and compute that $\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - \frac{1}{2}\mathbf{v}_1 = \begin{bmatrix} 3 \\ 3/2 \\ 3/2 \end{bmatrix}$. Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3/2 \\ 3/2 \end{bmatrix} \right\}$.

4. Set $\mathbf{v}_1 = \mathbf{x}_1$ and compute that $\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - (-2)\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 3 \end{bmatrix}$. Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 3 \end{bmatrix} \right\}$.

5. Set $\mathbf{v}_1 = \mathbf{x}_1$ and compute that $\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - 2\mathbf{v}_1 = \begin{bmatrix} 5 \\ 1 \\ -4 \\ -1 \end{bmatrix}$. Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} 1 \\ -4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ -4 \\ -1 \end{bmatrix} \right\}$.

6. Set $\mathbf{v}_1 = \mathbf{x}_1$ and compute that $\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - (-3)\mathbf{v}_1 = \begin{bmatrix} 4 \\ 6 \\ -3 \\ 0 \end{bmatrix}$. Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} 3 \\ -1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \\ -3 \\ 0 \end{bmatrix} \right\}$.

7. Since $\|\mathbf{v}_1\| = \sqrt{30}$ and $\|\mathbf{v}_2\| = \sqrt{27/2} = 3\sqrt{6}/2,$ an orthonormal basis for W is

$$\left\{\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}\right\} = \left\{\begin{bmatrix} 2/\sqrt{30} \\ -5/\sqrt{30} \\ 1/\sqrt{30} \end{bmatrix}, \begin{bmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}\right\}.$$

8. Since $\|\mathbf{v}_1\| = \sqrt{50}$ and $\|\mathbf{v}_2\| = \sqrt{54} = 3\sqrt{6},$ an orthonormal basis for W is

$$\left\{\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}\right\} = \left\{\begin{bmatrix} 3/\sqrt{50} \\ -4/\sqrt{50} \\ 5/\sqrt{50} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}\right\}.$$

9. Call the columns of the matrix $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ and perform the Gram-Schmidt process on these vectors:

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - (-2)\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ 3 \\ -1 \end{bmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \mathbf{x}_3 - \frac{3}{2}\mathbf{v}_1 - \left(-\frac{1}{2}\right)\mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 3 \end{bmatrix}$$

Thus an orthogonal basis for W is $\left\{\begin{bmatrix} 3 \\ 1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 1 \\ 3 \end{bmatrix}\right\}.$

10. Call the columns of the matrix $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ and perform the Gram-Schmidt process on these vectors:

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - (-3)\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \mathbf{x}_3 - \frac{1}{2}\mathbf{v}_1 - \frac{5}{2}\mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 3 \\ -1 \end{bmatrix}$$

Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} -1 \\ 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 3 \\ -1 \end{bmatrix} \right\}$.

11. Call the columns of the matrix $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ and perform the Gram-Schmidt process on these vectors:

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - (-1)\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ 3 \\ -3 \\ 3 \end{bmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \mathbf{x}_3 - 4\mathbf{v}_1 - \left(-\frac{1}{3}\right)\mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \\ -2 \end{bmatrix}$$

Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 3 \\ -3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \\ -2 \end{bmatrix} \right\}$.

12. Call the columns of the matrix $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ and perform the Gram-Schmidt process on these vectors:

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - 4\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \mathbf{x}_3 - \frac{7}{2}\mathbf{v}_1 - \frac{3}{2}\mathbf{v}_2 = \begin{bmatrix} 3 \\ 3 \\ 0 \\ -3 \\ 3 \end{bmatrix}$$

Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 0 \\ -3 \\ 3 \end{bmatrix} \right\}$.

13. Since A and Q are given, $R = Q^T A = \begin{bmatrix} 5/6 & 1/6 & -3/6 & 1/6 \\ -1/6 & 5/6 & 1/6 & 3/6 \end{bmatrix} \begin{bmatrix} 5 & 9 \\ 1 & 7 \\ -3 & -5 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 6 & 12 \\ 0 & 6 \end{bmatrix}$.

14. Since A and Q are given, $R = Q^T A = \begin{bmatrix} -2/7 & 5/7 & 2/7 & 4/7 \\ 5/7 & 2/7 & -4/7 & 2/7 \end{bmatrix} \begin{bmatrix} -2 & 3 \\ 5 & 7 \\ 2 & -2 \\ 4 & 6 \end{bmatrix} = \begin{bmatrix} 7 & 7 \\ 0 & 7 \end{bmatrix}$.

15. The columns of Q will be normalized versions of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ found in Exercise 11.

Thus $Q = \begin{bmatrix} 1/\sqrt{5} & 1/2 & 1/2 \\ -1/\sqrt{5} & 0 & 0 \\ -1/\sqrt{5} & 1/2 & 1/2 \\ 1/\sqrt{5} & -1/2 & 1/2 \\ 1/\sqrt{5} & 1/2 & -1/2 \end{bmatrix}$, $R = Q^T A = \begin{bmatrix} \sqrt{5} & -\sqrt{5} & 4\sqrt{5} \\ 0 & 6 & -2 \\ 0 & 0 & 4 \end{bmatrix}$.

16. The columns of Q will be normalized versions of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ found in Exercise 12.

Thus $Q = \begin{bmatrix} 1/2 & -1/(2\sqrt{2}) & 1/2 \\ -1/2 & 1/(2\sqrt{2}) & 1/2 \\ 0 & 1/\sqrt{2} & 0 \\ 1/2 & 1/(2\sqrt{2}) & -1/2 \\ 1/2 & 1/(2\sqrt{2}) & 1/2 \end{bmatrix}$, $R = Q^T A = \begin{bmatrix} 2 & 8 & 7 \\ 0 & 2\sqrt{2} & 3\sqrt{2} \\ 0 & 0 & 6 \end{bmatrix}$.

17. **a**. False. Scaling was used in Example 2, but the scale factor was nonzero.

b. True. See (1) in the statement of Theorem 11.

c. True. See the solution of Example 4.

18. **a**. False. The three orthogonal vectors must be *nonzero* to be a basis for a three-dimensional subspace. (This was the case in Step 3 of the solution of Example 2.)

b. True. If $\mathbf{x}$ is not in a subspace W, then $\mathbf{x}$ cannot equal $\text{proj}_W \mathbf{x}$, because $\text{proj}_W \mathbf{x}$ is in W. This idea was used for $\mathbf{v}_{k+1}$ in the proof of Theorem 11.

c. True. See Theorem 12.

19. Suppose that $\mathbf{x}$ satisfies $R\mathbf{x} = \mathbf{0}$; then $QR\mathbf{x} = Q\mathbf{0} = \mathbf{0}$, and $A\mathbf{x} = \mathbf{0}$. Since the columns of A are linearly independent, $\mathbf{x}$ must be $\mathbf{0}$. This fact, in turn, shows that the columns of R are linearly indepedent. Since R is square, it is invertible by the Invertible Matrix Theorem.

20. If $\mathbf{y}$ is in Col$A$, then $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x}$. Then $\mathbf{y} = QR\mathbf{x} = Q(R\mathbf{x})$, which shows that $\mathbf{y}$ is a linear combination of the columns of Q using the entries in $R\mathbf{x}$ as weights. Conversly, suppose that $\mathbf{y} = Q\mathbf{x}$ for some $\mathbf{x}$. Since R is invertible, the equation $A = QR$ implies that $Q = AR^{-1}$. So $\mathbf{y} = AR^{-1}\mathbf{x} = A(R^{-1}\mathbf{x})$, which shows that $\mathbf{y}$ is in Col A.

21. Denote the columns of Q by $\{\mathbf{q}_1,\ldots,\mathbf{q}_n\}$. Note that $n \le m$, because A is $m \times n$ and has linearly independent columns. The columns of Q can be extended to an orthonormal basis for $\mathbb{R}^m$ as follows. Let $\mathbf{f}_1$ be the first vector in the standard basis for $\mathbb{R}^m$ that is *not* in $W_n = \text{Span}\{\mathbf{q}_1,\ldots,\mathbf{q}_n\}$, let $\mathbf{u}_1 = \mathbf{f}_1 - \text{proj}_{W_n}\mathbf{f}_1$, and let $\mathbf{q}_{n+1} = \mathbf{u}_1 / \|\mathbf{u}_1\|$. Then $\{\mathbf{q}_1,\ldots,\mathbf{q}_n,\mathbf{q}_{n+1}\}$ is an orthonormal basis for $W_{n+1} = \text{Span}\{\mathbf{q}_1,\ldots,\mathbf{q}_n,\mathbf{q}_{n+1}\}$. Next let $\mathbf{f}_2$ be the first vector in the standard basis for $\mathbb{R}^m$ that is *not* in W_{n+1}, let $\mathbf{u}_2 = \mathbf{f}_2 - \text{proj}_{W_{n+1}}\mathbf{f}_2$, and let $\mathbf{q}_{n+2} = \mathbf{u}_2 / \|\mathbf{u}_2\|$. Then $\{\mathbf{q}_1,\ldots,\mathbf{q}_n,\mathbf{q}_{n+1},\mathbf{q}_{n+2}\}$ is an orthogonal basis for $W_{n+2} = \text{Span}\{\mathbf{q}_1,\ldots,\mathbf{q}_n,\mathbf{q}_{n+1},\mathbf{q}_{n+2}\}$. This process will continue until $m - n$ vectors have been added to the original n vectors, and $\{\mathbf{q}_1,\ldots,\mathbf{q}_n,\mathbf{q}_{n+1},\ldots,\mathbf{q}_m\}$ is an orthonormal basis for $\mathbb{R}^n$. Let $Q_0 = [\mathbf{q}_{n+1} \quad \cdots \quad \mathbf{q}_m]$ and $Q_1 = [Q \quad Q_0]$. Then, using partitioned matrix multiplication, $Q_1\begin{bmatrix} R \\ O \end{bmatrix} = QR = A$.

22. We may assume that $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ is an orthonormal basis for W, by normalizing the vectors in the original basis given for W, if necessary. Let U be the matrix whose columns are $\mathbf{u}_1,\ldots,\mathbf{u}_p$. Then, by Theorem 10 in Section 6.3, $T(\mathbf{x}) = \text{proj}_W\mathbf{x} = (UU^T)\mathbf{x}$ for $\mathbf{x}$ in $\mathbb{R}^n$. Thus T is a matrix transformation and hence is a linear transformation, as was shown in Section 1.8.

23. Given $A = QR$, partition $A = [A_1 \quad A_2]$, where A_1 has p columns. Partition Q as $Q = [Q_1 \quad Q_2]$ where Q_1 has p columns, and partition R as $R = \begin{bmatrix} R_{11} & R_{12} \\ O & R_{22} \end{bmatrix}$, where R_{11} is a $p \times p$ matrix. Then

$$A = [A_1 \quad A_2] = QR = [Q_1 \quad Q_2]\begin{bmatrix} R_{11} & R_{12} \\ O & R_{22} \end{bmatrix} = [Q_1R_{11} \quad Q_1R_{12} + Q_2R_{22}]$$

Thus $A_1 = Q_1R_{11}$. The matrix Q_1 has orthonormal columns because its columns come from Q. The matrix R_{11} is square and upper triangular due to its position within the upper triangular matrix R. The diagonal entries of R_{11} are positive because they are diagonal entries of R. Thus Q_1R_{11} is a QR factorization of A_1.

24. **[M]** Call the columns of the matrix $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, and $\mathbf{x}_4$ and perform the Gram-Schmidt process on these vectors:

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - (-1)\mathbf{v}_1 = \begin{bmatrix} 3 \\ 3 \\ -3 \\ 0 \\ 3 \end{bmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \mathbf{x}_3 - \left(-\frac{1}{2}\right)\mathbf{v}_1 - \left(-\frac{4}{3}\right)\mathbf{v}_2 = \begin{bmatrix} 6 \\ 0 \\ 6 \\ 6 \\ 0 \end{bmatrix}$$

$$\mathbf{v}_4 = \mathbf{x}_4 - \frac{\mathbf{x}_4 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_4 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \frac{\mathbf{x}_4 \cdot \mathbf{v}_3}{\mathbf{v}_3 \cdot \mathbf{v}_3}\mathbf{v}_3 = \mathbf{x}_4 - \frac{1}{2}\mathbf{v}_1 - (-1)\mathbf{v}_2 - \left(-\frac{1}{2}\right)\mathbf{v}_3 = \begin{bmatrix} 0 \\ 5 \\ 0 \\ 0 \\ -5 \end{bmatrix}$$

Thus an orthogonal basis for W is $\left\{ \begin{bmatrix} -10 \\ 2 \\ -6 \\ 16 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ -3 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 0 \\ 6 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ 0 \\ 0 \\ -5 \end{bmatrix} \right\}$.

25. **[M]** The columns of Q will be normalized versions of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ found in Exercise 24. Thus

$$Q = \begin{bmatrix} -1/2 & 1/2 & 1/\sqrt{3} & 0 \\ 1/10 & 1/2 & 0 & 1/\sqrt{2} \\ -3/10 & -1/2 & 1/\sqrt{3} & 0 \\ 4/5 & 0 & 1/\sqrt{3} & 0 \\ 1/10 & 1/2 & 0 & -1/\sqrt{2} \end{bmatrix}, R = Q^T A = \begin{bmatrix} 20 & -20 & -10 & 10 \\ 0 & 6 & -8 & -6 \\ 0 & 0 & 6\sqrt{3} & -3\sqrt{3} \\ 0 & 0 & 0 & 5\sqrt{2} \end{bmatrix}$$

26. **[M]** In MATLAB, when A has n columns, suitable commands are

```
Q = A(:,1)/norm(A(:,1))
    %  The first column of Q
    for j=2: n
       v=A(:,j) – Q*(Q'*A(:,j))
       Q(:,j)=v/norm(v)
       % Add a new column to Q
    end
```

6.5 SOLUTIONS

Notes: This is a core section – the basic geometric principles in this section provide the foundation for all the applications in Sections 6.6–6.8. Yet this section need not take a full day. Each example provides a stopping place. Theorem 13 and Example 1 are all that is needed for Section 6.6. Theorem 15, however, gives an illustration of why the QR factorization is important. Example 4 is related to Exercise 17 in Section 6.6.

1. To find the normal equations and to find $\hat{\mathbf{x}}$, compute

$$A^T A = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -3 & 3 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & -11 \\ -11 & 22 \end{bmatrix}; \ A^T \mathbf{b} = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -3 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -4 \\ 11 \end{bmatrix}.$$

a. The normal equations are $(A^T A)\mathbf{x} = A^T \mathbf{b}$: $\begin{bmatrix} 6 & -11 \\ -11 & 22 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -4 \\ 11 \end{bmatrix}$.

b. Compute $\hat{x} = (A^T A)^{-1} A^T \mathbf{b} = \begin{bmatrix} 6 & -11 \\ -11 & 22 \end{bmatrix}^{-1} \begin{bmatrix} -4 \\ 11 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 22 & 11 \\ 11 & 6 \end{bmatrix} \begin{bmatrix} -4 \\ 11 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 33 \\ 22 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

2. To find the normal equations and to find $\hat{\mathbf{x}}$, compute

$$A^T A = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 12 & 8 \\ 8 & 10 \end{bmatrix}; \ A^T \mathbf{b} = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -5 \\ 8 \\ 1 \end{bmatrix} = \begin{bmatrix} -24 \\ -2 \end{bmatrix}.$$

a. The normal equations are $(A^T A)\mathbf{x} = A^T \mathbf{b}$: $\begin{bmatrix} 12 & 8 \\ 8 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -24 \\ -2 \end{bmatrix}$.

b. Compute $\hat{x} = (A^T A)^{-1} A^T \mathbf{b} = \begin{bmatrix} 12 & 8 \\ 8 & 10 \end{bmatrix}^{-1} \begin{bmatrix} -24 \\ -2 \end{bmatrix} = \frac{1}{56} \begin{bmatrix} 10 & -8 \\ -8 & 12 \end{bmatrix} \begin{bmatrix} -24 \\ -2 \end{bmatrix} = \frac{1}{56} \begin{bmatrix} -224 \\ 168 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$.

3. To find the normal equations and to find $\hat{\mathbf{x}}$, compute

$$A^T A = \begin{bmatrix} 1 & -1 & 0 & 2 \\ -2 & 2 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 6 & 42 \end{bmatrix}; \ A^T \mathbf{b} = \begin{bmatrix} 1 & -1 & 0 & 2 \\ -2 & 2 & 3 & 5 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \end{bmatrix}.$$

a. The normal equations are $(A^T A)\mathbf{x} = A^T \mathbf{b}$: $\begin{bmatrix} 6 & 6 \\ 6 & 42 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \end{bmatrix}$.

b. Compute $\hat{\mathbf{x}} = (\mathrm{A}^T \mathrm{A})^{-1} \mathrm{A}^T \mathbf{b} = \begin{bmatrix} 6 & 6 \\ 6 & 42 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ -6 \end{bmatrix} = \frac{1}{216} \begin{bmatrix} 42 & -6 \\ -6 & 6 \end{bmatrix} \begin{bmatrix} 6 \\ -6 \end{bmatrix}$.

$$= \frac{1}{216} \begin{bmatrix} 288 \\ -72 \end{bmatrix} = \begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix}$$

4. To find the normal equations and to find $\hat{\mathbf{x}}$, compute

$$A^TA=\begin{bmatrix}1&1&1\\3&-1&1\end{bmatrix}\begin{bmatrix}1&3\\1&-1\\1&1\end{bmatrix}=\begin{bmatrix}3&3\\3&11\end{bmatrix};\ A^T\mathbf{b}=\begin{bmatrix}1&1&1\\3&-1&1\end{bmatrix}\begin{bmatrix}5\\1\\0\end{bmatrix}=\begin{bmatrix}6\\14\end{bmatrix}.$$

a. The normal equations are $(A^TA)\mathbf{x}=A^T\mathbf{b}$: $\begin{bmatrix}3&3\\3&11\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}6\\14\end{bmatrix}$.

b. Compute $\hat{\mathbf{x}}=(\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}^T\mathbf{b}=\begin{bmatrix}3&3\\3&11\end{bmatrix}^{-1}\begin{bmatrix}6\\14\end{bmatrix}=\frac{1}{24}\begin{bmatrix}11&-3\\-3&3\end{bmatrix}\begin{bmatrix}6\\14\end{bmatrix}=\frac{1}{24}\begin{bmatrix}24\\24\end{bmatrix}=\begin{bmatrix}1\\1\end{bmatrix}$.

5. To find the least squares solutions to $A\mathbf{x}=\mathbf{b}$, compute and row reduce the augmented matrix for the system $A^TA\mathbf{x}=A^T\mathbf{b}$: $\begin{bmatrix}A^TA & A^T\mathbf{b}\end{bmatrix}=\begin{bmatrix}4&2&2&14\\2&2&0&4\\2&0&2&10\end{bmatrix}\sim\begin{bmatrix}1&0&1&5\\0&1&-1&-3\\0&0&0&0\end{bmatrix}$, so all vectors of the form $\hat{\mathbf{x}}=\begin{bmatrix}5\\-3\\0\end{bmatrix}+x_3\begin{bmatrix}-1\\1\\1\end{bmatrix}$ are the least-squares solutions of $A\mathbf{x}=\mathbf{b}$.

6. To find the least squares solutions to $A\mathbf{x}=\mathbf{b}$, compute and row reduce the augmented matrix for the system $A^TA\mathbf{x}=A^T\mathbf{b}$: $\begin{bmatrix}A^TA & A^T\mathbf{b}\end{bmatrix}=\begin{bmatrix}6&3&3&27\\3&3&0&12\\3&0&3&15\end{bmatrix}\sim\begin{bmatrix}1&0&1&5\\0&1&-1&-1\\0&0&0&0\end{bmatrix}$, so all vectors of the form $\hat{\mathbf{x}}=\begin{bmatrix}5\\-1\\0\end{bmatrix}+x_3\begin{bmatrix}-1\\1\\1\end{bmatrix}$ are the least-squares solutions of $A\mathbf{x}=\mathbf{b}$.

7. From Exercise 3, $A=\begin{bmatrix}1&-2\\-1&2\\0&3\\2&5\end{bmatrix}$, $\mathbf{b}=\begin{bmatrix}3\\1\\-4\\2\end{bmatrix}$, and $\hat{\mathbf{x}}=\begin{bmatrix}4/3\\-1/3\end{bmatrix}$. Since

$$A\hat{\mathbf{x}}-\mathbf{b}=\begin{bmatrix}1&-2\\-1&2\\0&3\\2&5\end{bmatrix}\begin{bmatrix}4/3\\-1/3\end{bmatrix}-\begin{bmatrix}3\\1\\-4\\2\end{bmatrix}=\begin{bmatrix}2\\-2\\-1\\1\end{bmatrix}-\begin{bmatrix}3\\1\\-4\\2\end{bmatrix}=\begin{bmatrix}-1\\-3\\3\\-1\end{bmatrix}$$, the least squares error is

$\|A\hat{\mathbf{x}}-\mathbf{b}\|=\sqrt{20}=2\sqrt{5}.$

8. From Exercise 4, $A=\begin{bmatrix}1&3\\1&-1\\1&1\end{bmatrix}$, $\mathbf{b}=\begin{bmatrix}5\\1\\0\end{bmatrix}$, and $\hat{\mathbf{x}}=\begin{bmatrix}1\\1\end{bmatrix}$. Since

$$A\hat{\mathbf{x}}-\mathbf{b}=\begin{bmatrix}1&3\\1&-1\\1&1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}-\begin{bmatrix}5\\1\\0\end{bmatrix}=\begin{bmatrix}4\\0\\2\end{bmatrix}-\begin{bmatrix}5\\1\\0\end{bmatrix}=\begin{bmatrix}-1\\-1\\2\end{bmatrix}$$, the least squares error is $\|A\hat{\mathbf{x}}-\mathbf{b}\|=\sqrt{6}$.

9. a. Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the method of Example 4 may be used to find $\hat{\mathbf{b}}$, the orthogonal projection of $\mathbf{b}$ onto Col A:

$$\hat{\mathbf{b}}=\frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1+\frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2=\frac{2}{7}\mathbf{a}_1+\frac{1}{7}\mathbf{a}_2=\frac{2}{7}\begin{bmatrix}1\\3\\-2\end{bmatrix}+\frac{1}{7}\begin{bmatrix}5\\1\\4\end{bmatrix}=\begin{bmatrix}1\\1\\0\end{bmatrix}.$$

b. The vector $\hat{\mathbf{x}}$ contains the weights which must be placed on $\mathbf{a}_1$ and $\mathbf{a}_2$ to produce $\hat{\mathbf{b}}$. These weights are easily read from the above equation, so $\hat{\mathbf{x}}=\begin{bmatrix}2/7\\1/7\end{bmatrix}$.

10. a. Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the method of Example 4 may be used to find $\hat{\mathbf{b}}$, the orthogonal projection of $\mathbf{b}$ onto Col A:

$$\hat{\mathbf{b}}=\frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1+\frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2=3\mathbf{a}_1+\frac{1}{2}\mathbf{a}_2=3\begin{bmatrix}1\\-1\\1\end{bmatrix}+\frac{1}{2}\begin{bmatrix}2\\4\\2\end{bmatrix}=\begin{bmatrix}4\\-1\\4\end{bmatrix}.$$

b. The vector $\hat{\mathbf{x}}$ contains the weights which must be placed on $\mathbf{a}_1$ and $\mathbf{a}_2$ to produce $\hat{\mathbf{b}}$. These weights are easily read from the above equation, so $\hat{\mathbf{x}}=\begin{bmatrix}3\\1/2\end{bmatrix}$.

11. a. Because the columns $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ of A are orthogonal, the method of Example 4 may be used to find $\hat{\mathbf{b}}$, the orthogonal projection of $\mathbf{b}$ onto Col A:

$$\hat{\mathbf{b}}=\frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1+\frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2+\frac{\mathbf{b}\cdot\mathbf{a}_3}{\mathbf{a}_3\cdot\mathbf{a}_3}\mathbf{a}_3=\frac{2}{3}\mathbf{a}_1+0\mathbf{a}_2+\frac{1}{3}\mathbf{a}_3$$

$$=\frac{2}{3}\begin{bmatrix}4\\1\\6\\1\end{bmatrix}+0\begin{bmatrix}0\\-5\\1\\-1\end{bmatrix}+\frac{1}{3}\begin{bmatrix}1\\1\\0\\-5\end{bmatrix}=\begin{bmatrix}3\\1\\4\\-1\end{bmatrix}.$$

b. The vector $\hat{\mathbf{x}}$ contains the weights which must be placed on $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ to produce $\hat{\mathbf{b}}$. These weights are easily read from the above equation, so $\hat{\mathbf{x}}=\begin{bmatrix}2/3\\0\\1/3\end{bmatrix}$.

12. a. Because the columns $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ of A are orthogonal, the method of Example 4 may be used to find $\hat{\mathbf{b}}$, the orthogonal projection of $\mathbf{b}$ onto Col A:

$$\hat{\mathbf{b}} = \frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2 + \frac{\mathbf{b}\cdot\mathbf{a}_3}{\mathbf{a}_3\cdot\mathbf{a}_3}\mathbf{a}_3 = \frac{1}{3}\mathbf{a}_1 + \frac{14}{3}\mathbf{a}_2 + \left(-\frac{5}{3}\right)\mathbf{a}_3$$

$$= \frac{1}{3}\begin{bmatrix}1\\1\\0\\-1\end{bmatrix} + \frac{14}{3}\begin{bmatrix}1\\0\\1\\1\end{bmatrix} - \frac{5}{3}\begin{bmatrix}0\\-1\\1\\-1\end{bmatrix} = \begin{bmatrix}5\\2\\3\\6\end{bmatrix}$$

b. The vector $\hat{\mathbf{x}}$ contains the weights which must be placed on $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ to produce $\hat{\mathbf{b}}$. These weights are easily read from the above equation, so $\hat{\mathbf{x}} = \begin{bmatrix}1/3\\14/3\\-5/3\end{bmatrix}$.

13. One computes that $A\mathbf{u} = \begin{bmatrix}11\\-11\\11\end{bmatrix}, \mathbf{b} - A\mathbf{u} = \begin{bmatrix}0\\2\\-6\end{bmatrix}, \|\mathbf{b} - A\mathbf{u}\| = \sqrt{40}$;

$$A\mathbf{v} = \begin{bmatrix}7\\-12\\7\end{bmatrix}, \mathbf{b} - A\mathbf{v} = \begin{bmatrix}4\\3\\-2\end{bmatrix}, \|\mathbf{b} - A\mathbf{v}\| = \sqrt{29}.$$

Since $A\mathbf{v}$ is closer to $\mathbf{b}$ than $A\mathbf{u}$ is, $A\mathbf{u}$ is not the closest point in Col A to $\mathbf{b}$. Thus $\mathbf{u}$ cannot be a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

14. One computes that $A\mathbf{u} = \begin{bmatrix}3\\8\\2\end{bmatrix}, \mathbf{b} - A\mathbf{u} = \begin{bmatrix}2\\-4\\2\end{bmatrix}, \|\mathbf{b} - A\mathbf{u}\| = \sqrt{24}$;

$$A\mathbf{v} = \begin{bmatrix}7\\2\\8\end{bmatrix}, \mathbf{b} - A\mathbf{v} = \begin{bmatrix}-2\\2\\-4\end{bmatrix}, \|\mathbf{b} - A\mathbf{v}\| = \sqrt{24}.$$

Since $A\mathbf{u}$ and $A\mathbf{v}$ are equally close to $\mathbf{b}$, and the orthogonal projection is the *unique* closest point in Col A to $\mathbf{b}$, neither $A\mathbf{u}$ nor $A\mathbf{v}$ can be the closest point in Col A to $\mathbf{b}$. Thus neither $\mathbf{u}$ nor $\mathbf{v}$ can be a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

15. The least squares solution satisfies $R\hat{\mathbf{x}} = Q^T\mathbf{b}$. Since $R = \begin{bmatrix}3 & 5\\0 & 1\end{bmatrix}$ and $Q^T\mathbf{b} = \begin{bmatrix}7\\-1\end{bmatrix}$, the augmented matrix for the system may be row reduced to find $\begin{bmatrix}R & Q^T\mathbf{b}\end{bmatrix} = \begin{bmatrix}3 & 5 & 7\\0 & 1 & -1\end{bmatrix} \sim \begin{bmatrix}1 & 0 & 4\\0 & 1 & -1\end{bmatrix}$ and so $\hat{\mathbf{x}} = \begin{bmatrix}4\\-1\end{bmatrix}$ is the least squares solution of $A\mathbf{x} = \mathbf{b}$.

16. The least squares solution satisfies $R\hat{\mathbf{x}} = Q^T\mathbf{b}$. Since $R = \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}$ and $Q^T\mathbf{b} = \begin{bmatrix} 17/2 \\ 9/2 \end{bmatrix}$, the augmented matrix for the system may be row reduced to find $\begin{bmatrix} R & Q^T\mathbf{b} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 17/2 \\ 0 & 5 & 9/2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2.9 \\ 0 & 1 & .9 \end{bmatrix}$, and so $\hat{\mathbf{x}} = \begin{bmatrix} 2.9 \\ .9 \end{bmatrix}$ is the least squares solution of $A\mathbf{x} = \mathbf{b}$.

17. **a.** True. See the beginning of the section. The distance from $A\mathbf{x}$ to $\mathbf{b}$ is $\| A\mathbf{x} - \mathbf{b} \|$.
b. True. See the comments about equation (1).
c. False. The inequality points in the wrong direction. See the definition of a least-squares solution.
d. True. See Theorem 13.
e. True. See Theorem 14.

18. **a.** True. See the paragraph following the definition of a least-squares solution.
b. False. If $\hat{\mathbf{x}}$ is the least-squares solution, then $A\hat{\mathbf{x}}$ is the point in the column space of A closest to $\mathbf{b}$. See Figure 1 and the paragraph preceding it.
c. True. See the discussion following equation (1).
d. False. The formula applies only when the columns of A are linearly independent. See Theorem 14.
e. False. See the comments after Example 4.
f. False. See the Numerical Note.

19. **a.** If $A\mathbf{x} = \mathbf{0}$, then $A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$. This shows that Nul A is contained in Nul A^TA.
b. If $A^TA\mathbf{x} = \mathbf{0}$, then $\mathbf{x}^TA^TA\mathbf{x} = \mathbf{x}^T\mathbf{0} = 0$. So $(A\mathbf{x})^T(A\mathbf{x}) = 0$, which means that $\| A\mathbf{x} \|^2 = 0$, and hence $A\mathbf{x} = \mathbf{0}$. This shows that Nul A^TA is contained in Nul A.

20. Suppose that $A\mathbf{x} = \mathbf{0}$. Then $A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$. Since A^TA is invertible, $\mathbf{x}$ must be $\mathbf{0}$. Hence the columns of A are linearly independent.

21. **a.** If A has linearly independent columns, then the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. By Exercise 19, the equation $A^TA\mathbf{x} = \mathbf{0}$ also has only the trivial solution. Since A^TA is a square matrix, it must be invertible by the Invertible Matrix Theorem.
b. Since the n linearly independent columns of A belong to $\mathbb{R}^m$, m could not be less than n.
c. The n linearly independent columns of A form a basis for Col A, so the rank of A is n.

22. Note that A^TA has n columns because A does. Then by the Rank Theorem and Exercise 19,

$$\operatorname{rank} A^TA = n - \dim \operatorname{Nul} A^TA = n - \dim \operatorname{Nul} A = \operatorname{rank} A$$

23. By Theorem 14, $\hat{\mathbf{b}} = A\hat{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}$. The matrix $A(A^TA)^{-1}A^T$ is sometimes called the *hat-matrix* in statistics.

24. Since in this case $A^TA = I$, the normal equations give $\hat{\mathbf{x}} = A^T\mathbf{b}$.

25. The normal equations are $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$, whose solution is the set of all (x, y) such that $x + y = 3$. The solutions correspond to the points on the line midway between the lines $x + y = 2$ and $x + y = 4$.

26. **[M]** Using .7 as an approximation for $\sqrt{2}/2$, $a_0 = a_2 \approx .353535$ and $a_1 = .5$. Using .707 as an approximation for $\sqrt{2}/2$, $a_0 = a_2 \approx .35355339$, $a_1 = .5$.

6.6 SOLUTIONS

Notes: This section is a valuable reference for any person who works with data that requires statistical analysis. Many graduate fields require such work. Science students in particular will benefit from Example 1. The general linear model and the subsequent examples are aimed at students who may take a multivariate statistics course. That may include more students than one might expect.

1. The design matrix X and the observation vector **y** are $X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}$, and one can compute

$X^T X = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 6 \\ 11 \end{bmatrix}, \hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T\mathbf{y} = \begin{bmatrix} .9 \\ .4 \end{bmatrix}$. The least-squares line $y = \beta_0 + \beta_1 x$ is thus $y = .9 + .4x$.

2. The design matrix X and the observation vector **y** are $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$, and one can compute

$X^T X = \begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 6 \\ 25 \end{bmatrix}, \hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T\mathbf{y} = \begin{bmatrix} -.6 \\ .7 \end{bmatrix}$. The least-squares line $y = \beta_0 + \beta_1 x$ is thus $y = -.6 + .7x$.

3. The design matrix X and the observation vector $\boldsymbol{y}$ are $X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix}$,

and one can compute $X^T X = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}, \hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T\mathbf{y} = \begin{bmatrix} 1.1 \\ 1.3 \end{bmatrix}$. The least-squares line $y = \beta_0 + \beta_1 x$ is thus $y = 1.1 + 1.3x$.

4. The design matrix X and the observation vector **y** are $X = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 6 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 3 \\ 2 \\ 1 \\ 0 \end{bmatrix}$, and one can compute

$X^T X = \begin{bmatrix} 4 & 16 \\ 16 & 74 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 6 \\ 17 \end{bmatrix}, \hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T\mathbf{y} = \begin{bmatrix} 4.3 \\ -.7 \end{bmatrix}$. The least-squares line $y = \beta_0 + \beta_1 x$ is thus $y = 4.3 - .7x$.

5. If two data points have different x-coordinates, then the two columns of the design matrix X cannot be multiples of each other and hence are linearly independent. By Theorem 14 in Section 6.5, the normal equations have a unique solution.

6. If the columns of X were linearly dependent, then the same dependence relation would hold for the vectors in $\mathbb{R}^3$ formed from the top three entries in each column. That is, the columns of the matrix $\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}$ would also be linearly dependent, and so this matrix (called a Vandermonde matrix) would be noninvertible. Note that the determinant of this matrix is $(x_2 - x_1)(x_3 - x_1)(x_3 - x_2) \neq 0$ since x_1, x_2, and x_3 are distinct. Thus this matrix is invertible, which means that the columns of X are in fact linearly independent. By Theorem 14 in Section 6.5, the normal equations have a unique solution.

7. **a**. The model that produces the correct least-squares fit is $\mathbf{y} = X\boldsymbol{\beta} + \epsilon$ where

$$X = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \\ 5 & 25 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1.8 \\ 2.7 \\ 3.4 \\ 3.8 \\ 3.9 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}.$$

b. **[M]** One computes that (to two decimal places) $\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.76 \\ -.20 \end{bmatrix}$, so the desired least-squares equation is $y = 1.76x - .20x^2$.

8. **a**. The model that produces the correct least-squares fit is $\mathbf{y} = X\boldsymbol{\beta} + \epsilon$ where

$$X = \begin{bmatrix} x_1 & x_1^2 & x_1^3 \\ \vdots & \vdots & \vdots \\ x_n & x_n^2 & x_n^3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

b. **[M]** For the given data, $X = \begin{bmatrix} 4 & 16 & 64 \\ 6 & 36 & 216 \\ 8 & 64 & 512 \\ 10 & 100 & 1000 \\ 12 & 144 & 1728 \\ 14 & 196 & 2744 \\ 16 & 256 & 4096 \\ 18 & 324 & 5832 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1.58 \\ 2.08 \\ 2.5 \\ 2.8 \\ 3.1 \\ 3.4 \\ 3.8 \\ 4.32 \end{bmatrix}$, so

$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y} = \begin{bmatrix} .5132 \\ -.03348 \\ .001016 \end{bmatrix}$, and the least-squares curve is

$y = .5132x - .03348x^2 + .001016x^3$.

9. The model that produces the correct least-squares fit is $\mathbf{y} = X\boldsymbol{\beta} + \epsilon$ where

$$X = \begin{bmatrix} \cos 1 & \sin 1 \\ \cos 2 & \sin 2 \\ \cos 3 & \sin 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 7.9 \\ 5.4 \\ -.9 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} A \\ B \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

10. **a**. The model that produces the correct least-squares fit is $\mathbf{y} = X\boldsymbol{\beta} + \epsilon$ where

$$X = \begin{bmatrix} e^{-.02(10)} & e^{-.07(10)} \\ e^{-.02(11)} & e^{-.07(11)} \\ e^{-.02(12)} & e^{-.07(12)} \\ e^{-.02(14)} & e^{-.07(14)} \\ e^{-.02(15)} & e^{-.07(15)} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 21.34 \\ 20.68 \\ 20.05 \\ 18.87 \\ 18.30 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} M_A \\ M_B \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix},$$

b. [M] One computes that (to two decimal places) $\hat{\boldsymbol{\beta}} = \begin{bmatrix} 19.94 \\ 10.10 \end{bmatrix}$, so the desired least-squares equation is $y = 19.94e^{-.02t} + 10.10e^{-.07t}$.

11. **[M]** The model that produces the correct least-squares fit is $\mathbf{y} = X\boldsymbol{\beta} + \epsilon$ where

$X = \begin{bmatrix} 1 & 3\cos .88 \\ 1 & 2.3\cos 1.1 \\ 1 & 1.65\cos 1.42 \\ 1 & 1.25\cos 1.77 \\ 1 & 1.01\cos 2.14 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 3 \\ 2.3 \\ 1.65 \\ 1.25 \\ 1.01 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta \\ e \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$. One computes that (to two decimal places) $\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.45 \\ .811 \end{bmatrix}$. Since $e = .811 < 1$ the orbit is an ellipse. The equation $r = \beta / (1 - e \cos \vartheta)$ produces $r = 1.33$ when $\vartheta = 4.6$.

12. **[M]** The model that produces the correct least-squares fit is $\mathbf{y} = X\boldsymbol{\beta} + \epsilon$, where

$$X = \begin{bmatrix} 1 & 3.78 \\ 1 & 4.11 \\ 1 & 4.39 \\ 1 & 4.73 \\ 1 & 4.88 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 91 \\ 98 \\ 103 \\ 110 \\ 112 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \text{ and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}.$$ One computes that (to two decimal places)

$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 18.56 \\ 19.24 \end{bmatrix}$, so the desired least-squares equation is $p = 18.56 + 19.24 \ln w$. When $w = 100$, $p \approx 107$ millimeters of mercury.

13. **[M]**

a. The model that produces the correct least-squares fit is $\mathbf{y} = X\boldsymbol{\beta} + \epsilon$ where

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 2^2 & 2^3 \\ 1 & 3 & 3^2 & 3^3 \\ 1 & 4 & 4^2 & 4^3 \\ 1 & 5 & 5^2 & 5^3 \\ 1 & 6 & 6^2 & 6^3 \\ 1 & 7 & 7^2 & 7^3 \\ 1 & 8 & 8^2 & 8^3 \\ 1 & 9 & 9^2 & 9^3 \\ 1 & 10 & 10^2 & 10^3 \\ 1 & 11 & 11^2 & 11^3 \\ 1 & 12 & 12^2 & 12^3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 8.8 \\ 29.9 \\ 62.0 \\ 104.7 \\ 159.1 \\ 222.0 \\ 294.5 \\ 380.4 \\ 471.1 \\ 571.7 \\ 686.8 \\ 809.2 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \text{ and } \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \end{bmatrix}.$$ One computes that (to four

decimal places) $\hat{\boldsymbol{\beta}} = \begin{bmatrix} -.8558 \\ 4.7025 \\ 5.5554 \\ -.0274 \end{bmatrix}$, so the desired least-squares polynomial is

$$y(t) = -.8558 + 4.7025t + 5.5554t^2 - .0274t^3.$$

b. The velocity $v(t)$ is the derivative of the position function $y(t)$, so
$v(t) = 4.7025 + 11.1108t - .0822t^2$, and $v(4.5) = 53.0$ ft/sec.

14. Write the design matrix as $[\mathbf{1} \quad \mathbf{x}]$. Since the residual vector $\epsilon = \mathbf{y} - X\hat{\boldsymbol{\beta}}$ is orthogonal to Col X,

$$0 = \mathbf{1} \cdot \epsilon = \mathbf{1} \cdot (\mathbf{y} - X\hat{\boldsymbol{\beta}}) = \mathbf{1}^T\mathbf{y} - (\mathbf{1}^T X)\hat{\boldsymbol{\beta}}$$

$$= (y_1 + \ldots + y_n) - \begin{bmatrix} n & \sum x \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \sum y - n\hat{\beta}_0 - \hat{\beta}_1 \sum x = n\overline{y} - n\hat{\beta}_0 - n\hat{\beta}_1\overline{x}$$

This equation may be solved for $\overline{y}$ to find $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1\overline{x}$.

15. Notice $X^TX = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$; $X^T\mathbf{y} = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$. The equations (7) in the text follow immediately from the normal equations $X^TX\boldsymbol{\beta} = X^T\mathbf{y}$.

16. The determinant of the coefficient matrix of the equations in (7) is $n\sum x^2 - (\sum x)^2$. Using the 2×2 formula for the inverse of the coefficient matrix, $\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \dfrac{1}{n\sum x^2 - (\sum x)^2}\begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{bmatrix}\begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$.

Hence $\hat{\beta}_0 = \dfrac{(\sum x^2)(\sum y) - (\sum x)(\sum xy)}{n\sum x^2 - (\sum x)^2}$, $\hat{\beta}_1 = \dfrac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$.

Note: A simple algebraic calculation shows that $\sum y - (\sum x)\hat{\beta}_1 = n\hat{\beta}_0$, which provides a simple formula for $\hat{\beta}_0$ once $\hat{\beta}_1$ is known

17. **a**. The mean of the data in Example 1 is $\bar{x} = 5.5$, so the data in mean-deviation form are (–3.5, 1), (–.5, 2), (1.5, 3), (2.5, 3), and the associated design matrix is $X = \begin{bmatrix} 1 & -3.5 \\ 1 & -.5 \\ 1 & 1.5 \\ 1 & 2.5 \end{bmatrix}$. The columns of X are orthogonal because the entries in the second column sum to 0.

b. The normal equations are $X^TX\boldsymbol{\beta} = X^T\mathbf{y}$, or $\begin{bmatrix} 4 & 0 \\ 0 & 21 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 7.5 \end{bmatrix}$. One computes that

$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 9/4 \\ 5/14 \end{bmatrix}$, so the desired least-squares line is $y = (9/4) + (5/14)x^* = (9/4) + (5/14)(x - 5.5)$.

18. Since $X^TX = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$, X^TX is a diagonal matrix when $\sum x = 0$.

19. The residual vector $\epsilon = \mathbf{y} - X\hat{\boldsymbol{\beta}}$ is orthogonal to Col X, while $\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}}$ is in Col X. Since ϵ and $\hat{\mathbf{y}}$ are thus orthogonal, apply the Pythagorean Theorem to these vectors to obtain
$\text{SS(T)} = \|\mathbf{y}\|^2 = \|\hat{\mathbf{y}} + \epsilon\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\epsilon\|^2 = \|X\hat{\beta}\|^2 + \|\mathbf{y} - X\hat{\beta}\|^2 = \text{SS(R)} + \text{SS(E)}$.

20. Since $\hat{\beta}$ satisfies the normal equations, $X^TX\hat{\boldsymbol{\beta}} = X^T\mathbf{y}$, and
$\|X\hat{\boldsymbol{\beta}}\|^2 = (X\hat{\boldsymbol{\beta}})^T(X\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}^TX^TX\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^TX^T\mathbf{y}$. Since $\|X\hat{\boldsymbol{\beta}}\|^2 = \text{SS(R)}$ and $\mathbf{y}^T\mathbf{y} = \|\mathbf{y}\|^2 = \text{SS(T)}$, Exercise 19 shows that $\text{SS(E)} = \text{SS(T)} - \text{SS(R)} = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^TX^T\mathbf{y}$.

6.7 SOLUTIONS

Notes: The three types of inner products described here (in Examples 1, 2, and 7) are matched by examples in Section 6.8. It is possible to spend just one day on selected portions of both sections. Example 1 matches the weighted least squares in Section 6.8. Examples 2–6 are applied to trend analysis in Seciton 6.8. This material is aimed at students who have not had much calculus or who intend to take more than one course in statistics.

For students who have seen some calculus, Example 7 is needed to develop the Fourier series in Section 6.8. Example 8 is used to motivate the inner product on $C[a, b]$. The Cauchy-Schwarz and triangle inequalities are not used here, but they should be part of the training of every mathematics student.

1. The inner product is $\langle x, y\rangle = 4x_1y_1 + 5x_2y_2$. Let $\mathbf{x} = (1, 1)$, $\mathbf{y} = (5, -1)$.

a. Since $\|\mathbf{x}\|^2 = \langle x, x\rangle = 9$, $\|\mathbf{x}\| = 3$. Since $\|\mathbf{y}\|^2 = \langle y, y\rangle = 105$, $\|\mathbf{y}\| = \sqrt{105}$. Finally, $|\langle x, y\rangle|^2 = 15^2 = 225$.

b. A vector $\mathbf{z}$ is orthogonal to $\mathbf{y}$ if and only if $\langle x, y\rangle = 0$, that is, $20z_1 - 5z_2 = 0$, or $4z_1 = z_2$. Thus all multiples of $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ are orthogonal to $\mathbf{y}$.

2. The inner product is $\langle x, y\rangle = 4x_1y_1 + 5x_2y_2$. Let $\mathbf{x} = (3, -2)$, $\mathbf{y} = (-2, 1)$. Compute that $\|\mathbf{x}\|^2 = \langle x, x\rangle = 56$, $\|\mathbf{y}\|^2 = \langle y, y\rangle = 21$, $\|\mathbf{x}\|^2\|\mathbf{y}\|^2 = 56\cdot 21 = 1176$, $\langle x, y\rangle = -34$, and $|\langle x, y\rangle|^2 = 1156$. Thus $|\langle x, y\rangle|^2 \le \|\mathbf{x}\|^2\|\mathbf{y}\|^2$, as the Cauchy-Schwarz inequality predicts.

3. The inner product is $\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$, so $\langle 4+t, 5-4t^2\rangle = 3(1) + 4(5) + 5(1) = 28$.

4. The inner product is $\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$, so $\langle 3t-t^2, 3+2t^2\rangle = (-4)(5) + 0(3) + 2(5) = -10$.

5. The inner product is $\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$, so $\langle p, p\rangle = \langle 4+t, 4+t\rangle = 3^2 + 4^2 + 5^2 = 50$ and $\|p\| = \sqrt{\langle p, p\rangle} = \sqrt{50} = 5\sqrt{2}$. Likewise $\langle q, q\rangle = \langle 5-4t^2, 5-4t^2\rangle = 1^2 + 5^2 + 1^2 = 27$ and $\|q\| = \sqrt{\langle q, q\rangle} = \sqrt{27} = 3\sqrt{3}$.

6. The inner product is $\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$, so $\langle p, p\rangle = \langle 3t-t^2, 3t-t^2\rangle = (-4)^2 + 0^2 + 2^2 = 20$ and $\|p\| = \sqrt{\langle p, p\rangle} = \sqrt{20} = 2\sqrt{5}$. Likewise $\langle q, q\rangle = \langle 3+2t^2, 3+2t^2\rangle = 5^2 + 3^2 + 5^2 = 59$ and $\|q\| = \sqrt{\langle q, q\rangle} = \sqrt{59}$.

7. The orthogonal projection $\hat{q}$ of q onto the subspace spanned by p is

$$\hat{q} = \frac{\langle q, p\rangle}{\langle p, p\rangle} p = \frac{28}{50}(4+t) = \frac{56}{25} + \frac{14}{25}t.$$

8. The orthogonal projection $\hat{q}$ of q onto the subspace spanned by p is

$$\hat{q} = \frac{\langle q, p\rangle}{\langle p, p\rangle} p = -\frac{10}{20}(3t - t^2) = -\frac{3}{2}t + \frac{1}{2}t^2 .$$

9. The inner product is $\langle p, q\rangle = p(-3)q(-3) + p(-1)q(-1) + p(1)q(1) + p(3)q(3)$.

a. The orthogonal projection $\hat{p}_2$ of p_2 onto the subspace spanned by p_0 and p_1 is

$$\hat{p}_2 = \frac{\langle p_2, p_0\rangle}{\langle p_0, p_0\rangle} p_0 + \frac{\langle p_2, p_1\rangle}{\langle p_1, p_1\rangle} p_1 = \frac{20}{4}(1) + \frac{0}{20}t = 5 .$$

b. The vector $q = p_2 - \hat{p}_2 = t^2 - 5$ will be orthogonal to both p_0 and p_1 and $\{p_0, p_1, q\}$ will be an orthogonal basis for $\text{Span}\{p_0, p_1, p_2\}$. The vector of values for q at (–3, –1, 1, 3) is (4, –4, –4, 4), so scaling by 1/4 yields the new vector $q = (1/4)(t^2 - 5)$.

10. The best approximation to $p = t^3$ by vectors in $W = \text{Span}\{p_0, p_1, q\}$ will be

$$\hat{p} = \text{proj}_W\, p = \frac{\langle p, p_0\rangle}{\langle p_0, p_0\rangle} p_0 + \frac{\langle p, p_1\rangle}{\langle p_1, p_1\rangle} p_1 + \frac{\langle p, q\rangle}{\langle q, q\rangle} q = \frac{0}{4}(1) + \frac{164}{20}(t) + \frac{0}{4}\left(\frac{t^2 - 5}{4}\right) = \frac{41}{5}t .$$

11. The orthogonal projection of $p = t^3$ onto $W = \text{Span}\{p_0, p_1, p_2\}$ will be

$$\hat{p} = \text{proj}_W\, p = \frac{\langle p, p_0\rangle}{\langle p_0, p_0\rangle} p_0 + \frac{\langle p, p_1\rangle}{\langle p_1, p_1\rangle} p_1 + \frac{\langle p, p_2\rangle}{\langle p_2, p_2\rangle} p_2 = \frac{0}{5}(1) + \frac{34}{10}(t) + \frac{0}{14}(t^2 - 2) = \frac{17}{5}t .$$

12. Let $W = \text{Span}\{p_0, p_1, p_2\}$. The vector $p_3 = p - \text{proj}_W\, p = t^3 - (17/5)t$ will make $\{p_0, p_1, p_2, p_3\}$ an orthogonal basis for the subspace $\mathbb{P}_3$ of $\mathbb{P}_4$. The vector of values for p_3 at (–2, –1, 0, 1, 2) is (–6/5, 12/5, 0, –12/5, 6/5), so scaling by 5/6 yields the new vector $p_3 = (5/6)(t^3 - (17/5)t) = (5/6)t^3 - (17/6)t$.

13. Suppose that A is invertible and that $\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$ for $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$. Check each axiom in the definition of an inner product space, using the properties of the dot product.

i. $\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{u}) \cdot (A\mathbf{v}) = (A\mathbf{v}) \cdot (A\mathbf{u}) = \langle \mathbf{v}, \mathbf{u}\rangle$

ii. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w}\rangle = (A(\mathbf{u} + \mathbf{v})) \cdot (A\mathbf{w}) = (A\mathbf{u} + A\mathbf{v}) \cdot (A\mathbf{w}) = (A\mathbf{u}) \cdot (A\mathbf{w}) + (A\mathbf{v}) \cdot (A\mathbf{w}) = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$

iii. $\langle c\mathbf{u}, \mathbf{v}\rangle = (A(c\mathbf{u})) \cdot (A\mathbf{v}) = (c(A\mathbf{u})) \cdot (A\mathbf{v}) = c((A\mathbf{u}) \cdot (A\mathbf{v})) = c\langle \mathbf{u}, \mathbf{v}\rangle$

iv. $\langle \mathbf{u}, \mathbf{u}\rangle = (A\mathbf{u}) \cdot (A\mathbf{u}) = \| A\mathbf{u} \|^2 \ge 0$, and this quantity is zero if and only if the vector $A\mathbf{u}$ is $\mathbf{0}$. But $A\mathbf{u} = \mathbf{0}$ if and only $\mathbf{u} = \mathbf{0}$ because A is invertible.

14. Suppose that T is a one-to-one linear transformation from a vector space V into $\mathbb{R}^n$ and that $\langle \mathbf{u}, \mathbf{v}\rangle = T(\mathbf{u}) \cdot T(\mathbf{v})$ for $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$. Check each axiom in the definition of an inner product space, using the properties of the dot product and T. The linearity of T is used often in the following.

i. $\langle \mathbf{u}, \mathbf{v}\rangle = T(\mathbf{u}) \cdot T(\mathbf{v}) = T(\mathbf{v}) \cdot T(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u}\rangle$

ii. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w}\rangle = T(\mathbf{u} + \mathbf{v}) \cdot T(\mathbf{w}) = (T(\mathbf{u}) + T(\mathbf{v})) \cdot T(\mathbf{w}) = T(\mathbf{u}) \cdot T(\mathbf{w}) + T(\mathbf{v}) \cdot T(\mathbf{w}) = \langle \mathbf{u}, w\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$

iii. $\langle c\mathbf{u}, \mathbf{v}\rangle = T(c\mathbf{u}) \cdot T(\mathbf{v}) = (cT(\mathbf{u})) \cdot T(\mathbf{v}) = c(T(\mathbf{u}) \cdot T(\mathbf{v})) = c\langle \mathbf{u}, \mathbf{v}\rangle$

iv. $\langle \mathbf{u}, \mathbf{u}\rangle = T(\mathbf{u}) \cdot T(\mathbf{u}) = \| T(\mathbf{u}) \|^2 \ge 0$, and this quantity is zero if and only if $\mathbf{u} = \mathbf{0}$ since T is a one-to-one transformation.

15. Using Axioms 1 and 3, $\langle \mathbf{u}, c\mathbf{v}\rangle = \langle c\mathbf{v}, \mathbf{u}\rangle = c\langle \mathbf{v}, \mathbf{u}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$.

16. Using Axioms 1, 2 and 3,

$$\begin{aligned}\|\mathbf{u}-\mathbf{v}\|^2 &= \langle \mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}-\mathbf{v}\rangle - \langle \mathbf{v}, \mathbf{u}-\mathbf{v}\rangle \\ &= \langle \mathbf{u},\mathbf{u}\rangle - \langle \mathbf{u},\mathbf{v}\rangle - \langle \mathbf{v},\mathbf{u}\rangle + \langle \mathbf{v},\mathbf{v}\rangle = \langle \mathbf{u},\mathbf{u}\rangle - 2\langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{v}\rangle \\ &= \|\mathbf{u}\|^2 - 2\langle \mathbf{u},\mathbf{v}\rangle + \|\mathbf{v}\|^2\end{aligned}$$

Since $\{\mathbf{u}, \mathbf{v}\}$ is orthonormal, $\|\mathbf{u}\|^2 = \|\mathbf{v}\|^2 = 1$ and $\langle \mathbf{u}, \mathbf{v}\rangle = 0$. So $\|\mathbf{u}-\mathbf{v}\|^2 = 2$.

17. Following the method in Exercise 16,

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}+\mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}+\mathbf{v}\rangle \\ &= \langle \mathbf{u},\mathbf{u}\rangle + \langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{u}\rangle + \langle \mathbf{v},\mathbf{v}\rangle = \langle \mathbf{u},\mathbf{u}\rangle + 2\langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{v}\rangle \\ &= \|\mathbf{u}\|^2 + 2\langle \mathbf{u},\mathbf{v}\rangle + \|\mathbf{v}\|^2\end{aligned}$$

Subtracting these results, one finds that $\|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2 = 4\langle \mathbf{u},\mathbf{v}\rangle$, and dividing by 4 gives the desired identity.

18. In Exercises 16 and 17, it has been shown that $\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 - 2\langle \mathbf{u},\mathbf{v}\rangle + \|\mathbf{v}\|^2$ and $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\langle \mathbf{u},\mathbf{v}\rangle + \|\mathbf{v}\|^2$. Adding these two results gives $\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$.

19. let $\mathbf{u} = \begin{bmatrix}\sqrt{a}\\ \sqrt{b}\end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix}\sqrt{b}\\ \sqrt{a}\end{bmatrix}$. Then $\|\mathbf{u}\|^2 = a+b$, $\|\mathbf{v}\|^2 = a+b$, and $\langle \mathbf{u},\mathbf{v}\rangle = 2\sqrt{ab}$. Since a and b are nonnegative, $\|\mathbf{u}\| = \sqrt{a+b}$, $\|\mathbf{v}\| = \sqrt{a+b}$. Plugging these values into the Cauchy-Schwarz inequality gives $2\sqrt{ab} = |\langle \mathbf{u},\mathbf{v}\rangle| \le \|\mathbf{u}\|\|\mathbf{v}\| = \sqrt{a+b}\sqrt{a+b} = a+b$. Dividing both sides of this equation by 2 gives the desired inequality.

20. The Cauchy-Schwarz inequality may be altered by dividing both sides of the inequality by 2 and then squaring both sides of the inequality. The result is $\left(\frac{\langle \mathbf{u},\mathbf{v}\rangle}{2}\right)^2 \le \frac{\|\mathbf{u}\|^2\|\mathbf{v}\|^2}{4}$. Now let $\mathbf{u} = \begin{bmatrix}a\\ b\end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix}1\\ 1\end{bmatrix}$. Then $\|\mathbf{u}\|^2 = a^2 + b^2$, $\|\mathbf{v}\|^2 = 2$, and $\langle \mathbf{u}, \mathbf{v}\rangle = a + b$. Plugging these values into the inequality above yields the desired inequality.

21. The inner product is $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$. Let $f(t) = 1-3t^2$, $g(t) = t - t^3$. Then $\langle f, g\rangle = \int_0^1 (1-3t^2)(t-t^3)\,dt = \int_0^1 3t^5 - 4t^3 + t\,dt = 0$.

22. The inner product is $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$. Let $f(t) = 5t - 3$, $g(t) = t^3 - t^2$. Then $\langle f, g\rangle = \int_0^1 (5t-3)(t^3-t^2)dt = \int_0^1 5t^4 - 8t^3 + 3t^2 dt = 0$.

23. The inner product is $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$, so $\langle f, f\rangle = \int_0^1 (1-3t^2)^2\,dt = \int_0^1 9t^4 - 6t^2 + 1\,dt = 4/5$, and $\| f \| = \sqrt{\langle f, f\rangle} = 2/\sqrt{5}$.

24. The inner product is $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$, so $\langle g, g\rangle = \int_0^1 (t^3-t^2)^2\,dt = \int_0^1 t^6 - 2t^5 + t^4 dt = 1/105$, and $\| g \| = \sqrt{\langle g, g\rangle} = 1/\sqrt{105}$.

25. The inner product is $\langle f, g\rangle = \int_{-1}^1 f(t)g(t)dt$. Then 1 and t are orthogonal because $\langle 1, t\rangle = \int_{-1}^1 t\,dt = 0$. So 1 and t can be in an orthogonal basis for $\text{Span}\{1, t, t^2\}$. By the Gram-Schmidt process, the third basis element in the orthogonal basis can be $t^2 - \frac{\langle t^2, 1\rangle}{\langle 1, 1\rangle}1 - \frac{\langle t^2, t\rangle}{\langle t, t\rangle}t$. Since $\langle t^2, 1\rangle = \int_{-1}^1 t^2 dt = 2/3$, $\langle 1, 1\rangle = \int_{-1}^1 1\,dt = 2$, and $\langle t^2, t\rangle = \int_{-1}^1 t^3 dt = 0$, the third basis element can be written as $t^2 - (1/3)$. This element can be scaled by 3, which gives the orthogonal basis as $\{1, t, 3t^2 - 1\}$.

26. The inner product is $\langle f, g\rangle = \int_{-2}^2 f(t)g(t)dt$. Then 1 and t are orthogonal because $\langle 1, t\rangle = \int_{-2}^2 t\,dt = 0$. So 1 and t can be in an orthogonal basis for $\text{Span}\{1, t, t^2\}$. By the Gram-Schmidt process, the third basis element in the orthogonal basis can be $t^2 - \frac{\langle t^2, 1\rangle}{\langle 1, 1\rangle}1 - \frac{\langle t^2, t\rangle}{\langle t, t\rangle}t$. Since $\langle t^2, 1\rangle = \int_{-2}^2 t^2 dt = 16/3$, $\langle 1, 1\rangle = \int_{-2}^2 1\,dt = 4$, and $\langle t^2, t\rangle = \int_{-2}^2 t^3 dt = 0$, the third basis element can be written as $t^2 - (4/3)$. This element can be scaled by 3, which gives the orthogonal basis as $\{1, t, 3t^2 - 4\}$.

27. **[M]** The new orthogonal polynomials are multiples of $-17t + 5t^3$ and $72 - 155t^2 + 35t^4$. These polynomials may be scaled so that their values at –2, –1, 0, 1, and 2 are small integers.

28. **[M]** The orthogonal basis is $f_0(t) = 1$, $f_1(t) = \cos t$, $f_2(t) = \cos^2 t - (1/2) = (1/2)\cos 2t$, and $f_3(t) = \cos^3 t - (3/4)\cos t = (1/4)\cos 3t$.

6.8 SOLUTIONS

Notes: The connections between this section and Section 6.7 are described in the notes for that section. For my junior-senior class, I spend three days on the following topics: Theorems 13 and 15 in Section 6.5, plus Examples 1, 3, and 5; Example 1 in Section 6.6; Examples 2 and 3 in Section 6.7, with the motivation for the definite integral; and Fourier series in Section 6.8.

1. The weighting matrix W, design matrix X, parameter vector β, and observation vector $\mathbf{y}$ are:

$$W = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, X = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 4 \\ 4 \end{bmatrix}.$$

The design matrix X and the observation vector $\mathbf{y}$ are scaled by W:

$$WX = \begin{bmatrix} 1 & -2 \\ 2 & -2 \\ 2 & 0 \\ 2 & 2 \\ 1 & 2 \end{bmatrix}, W\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 4 \\ 8 \\ 4 \end{bmatrix}.$$

Further compute . $(WX)^T WX = \begin{bmatrix} 14 & 0 \\ 0 & 16 \end{bmatrix}, (WX)^T W\mathbf{y} = \begin{bmatrix} 28 \\ 24 \end{bmatrix}$ and find that

$\hat{\boldsymbol{\beta}} = ((WX)^T WX)^{-1}(WX)^T W\mathbf{y} = \begin{bmatrix} 1/14 & 0 \\ 0 & 1/16 \end{bmatrix}\begin{bmatrix} 28 \\ 24 \end{bmatrix} = \begin{bmatrix} 2 \\ 3/2 \end{bmatrix}$. Thus the weighted least-squares line is $y = 2 + (3/2)x$.

2. Let X be the original design matrix, and let $\mathbf{y}$ be the original observation vector. Let W be the weighting matrix for the first method. Then $2W$ is the weighting matrix for the second method. The weighted least-squares by the first method is equivalent to the ordinary least-squares for an equation whose normal equation is $(WX)^T WX\hat{\boldsymbol{\beta}} = (WX)^T W\mathbf{y}$, while the second method is equivalent to the ordinary least-squares for an equation whose normal equation is $(2WX)^T (2W)X\hat{\boldsymbol{\beta}} = (2WX)^T (2W)\mathbf{y}$. Since the second equation can be written as $4(WX)^T WX\hat{\beta} = 4(WX)^T W\mathbf{y}$, it has the same solutions as the first equation).

3. From Example 2 and the statement of the problem, $p_0(t) = 1$, $p_1(t) = t$, $p_2(t) = t^2 - 2$, $p_3(t) = (5/6)t^3 - (17/6)t$, and $g = (3, 5, 5, 4, 3)$. The cubic trend function for g is the orthogonal projection $\hat{p}$ of g onto the subspace spanned by p_0, p_1, p_2, and p_3:

$$\begin{aligned}\hat{p} &= \frac{\langle g, p_0\rangle}{\langle p_0, p_0\rangle}p_0 + \frac{\langle g, p_1\rangle}{\langle p_1, p_1\rangle}p_1 + \frac{\langle g, p_2\rangle}{\langle p_2, p_2\rangle}p_2 + \frac{\langle g, p_3\rangle}{\langle p_3, p_3\rangle}p_3 \\ &= \frac{20}{5}(1) + \frac{-1}{10}t + \frac{-7}{14}\left(t^2 - 2\right) + \frac{2}{10}\left(\frac{5}{6}t^3 - \frac{17}{6}t\right) \\ &= 4 - \frac{1}{10}t - \frac{1}{2}\left(t^2 - 2\right) + \frac{1}{5}\left(\frac{5}{6}t^3 - \frac{17}{6}t\right) = 5 - \frac{2}{3}t - \frac{1}{2}t^2 + \frac{1}{6}t^3\end{aligned}$$

This polynomial happens to fit the data exactly.

4. The inner product is $\langle p, q\rangle = p(-5)q(-5) + p(-3)q(-3) + p(-1)q(-1) + p(1)q(1) + p(3)q(3) + p(5)q(5)$.

a. Begin with the basis $\{1, t, t^2\}$ for $\mathbb{P}_2$. Since 1 and t are orthogonal, let $p_0(t) = 1$ and $p_1(t) = t$. Then the Gram-Schmidt process gives $p_2(t) = t^2 - \frac{\langle t^2, 1\rangle}{\langle 1, 1\rangle}1 - \frac{\langle t^2, t\rangle}{\langle t, t\rangle}t = t^2 - \frac{70}{6} = t^2 - \frac{35}{3}$. The vector of values for p_2 is (40/3, –8/3, –32/3, –32/3, –8/3, 40/3), so scaling by 3/8 yields the new function $p_2 = (3/8)(t^2 - (35/3)) = (3/8)t^2 - (35/8)$.

b. The data vector is $g = (1, 1, 4, 4, 6, 8)$. The quadratic trend function for g is the orthogonal projection $\hat{p}$ of g onto the subspace spanned by p_0, p_1 and p_2:

$$\hat{p} = \frac{\langle g, p_0 \rangle}{\langle p_0, p_0 \rangle} p_0 + \frac{\langle g, p_1 \rangle}{\langle p_1, p_1 \rangle} p_1 + \frac{\langle g, p_2 \rangle}{\langle p_2, p_2 \rangle} p_2 = \frac{24}{6}(1) + \frac{50}{70}t + \frac{6}{84}\left(\frac{3}{8}t^2 - \frac{35}{8}\right)$$
$$= 4 + \frac{5}{7}t + \frac{1}{14}\left(\frac{3}{8}t^2 - \frac{35}{8}\right) = \frac{59}{16} + \frac{5}{7}t + \frac{3}{112}t^2$$

5. The inner product is $\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)dt$. Let $m \neq n$. Then $\langle \sin mt, \sin nt\rangle = \int_0^{2\pi} \sin mt \sin nt \, dt = \frac{1}{2}\int_0^{2\pi} \cos((m-n)t) - \cos((m+n)t)dt = 0$. Thus $\sin mt$ and $\sin nt$ are orthogonal.

6. The inner product is $\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)dt$. Let m and n be positive integers. Then $\langle \sin mt, \cos nt\rangle = \int_0^{2\pi} \sin mt \cos nt \, dt = \frac{1}{2}\int_0^{2\pi} \sin((m+n)t) + \sin((m-n)t)dt = 0$. Thus $\sin mt$ and $\cos nt$ are orthogonal.

7. The inner product is $\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)dt$. Let k be a positive integer. Then

$\|\cos kt\|^2 = \langle \cos kt, \cos kt\rangle = \int_0^{2\pi} \cos^2 kt \, dt = \frac{1}{2}\int_0^{2\pi} 1 + \cos 2kt \, dt = \pi$ and

$\|\sin kt\|^2 = \langle \sin kt, \sin kt\rangle = \int_0^{2\pi} \sin^2 kt \, dt = \frac{1}{2}\int_0^{2\pi} 1 - \cos 2kt \, dt = \pi$.

8. Let $f(t) = t - 1$. The Fourier coefficients for f are: $\frac{a_0}{2} = \frac{1}{2}\frac{1}{\pi}\int_0^{2\pi} f(t)\, dt = \frac{1}{2\pi}\int_0^{2\pi} t - 1\, dt = -1 + \pi$ and for $k > 0$, $a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos kt \, dt = \frac{1}{\pi}\int_0^{2\pi} (t-1)\cos kt \, dt = 0$, and $b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt \, dt = \frac{1}{\pi}\int_0^{2\pi} (t-1)\sin kt \, dt = -\frac{2}{k}$. The third-order Fourier approximation to f is thus $\frac{a_0}{2} + b_1 \sin t + b_2 \sin 2t + b_3 \sin 3t = -1 + \pi - 2\sin t - \sin 2t - \frac{2}{3}\sin 3t$.

9. Let $f(t) = 2\pi - t$. The Fourier coefficients for f are: $\frac{a_0}{2} = \frac{1}{2}\frac{1}{\pi}\int_0^{2\pi} f(t)\, dt = \frac{1}{2\pi}\int_0^{2\pi} 2\pi - t\, dt = \pi$ and for $k > 0$, $a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos kt \, dt = \frac{1}{\pi}\int_0^{2\pi} (2\pi - t)\cos kt \, dt = 0$ and $b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt \, dt = \frac{1}{\pi}\int_0^{2\pi} (2\pi - t)\sin kt \, dt = \frac{2}{k}$. The third-order Fourier approximation to f is thus $\frac{a_0}{2} + b_1 \sin t + b_2 \sin 2t + b_3 \sin 3t = \pi + 2\sin t + \sin 2t + \frac{2}{3}\sin 3t$.

10. Let $f(t) = \begin{cases} 1 & \text{for } 0 \le t < \pi \\ -1 & \text{for } \pi \le t < 2\pi \end{cases}$. The Fourier coefficients for f are:

$\frac{a_0}{2} = \frac{1}{2}\frac{1}{\pi}\int_0^{2\pi} f(t)\, dt = \frac{1}{2\pi}\int_0^{\pi} dt - \frac{1}{2\pi}\int_\pi^{2\pi} dt = 0$, and for $k > 0$,

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos kt\,dt = \frac{1}{\pi}\int_0^{\pi}\cos kt\,dt - \frac{1}{\pi}\int_{\pi}^{2\pi}\cos kt\,dt = 0 \text{ and}$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt\,dt = \frac{1}{\pi}\int_0^{\pi}\sin kt\,dt - \frac{1}{\pi}\int_{\pi}^{2\pi}\sin kt\,dt = \begin{cases} 4/(k\pi) & \text{for } k \text{ odd} \\ 0 & \text{for } k \text{ even} \end{cases}.$$

The third-order Fourier approximation to f is thus $b_1 \sin t + b_3 \sin 3t = \frac{4}{\pi}\sin t + \frac{4}{3\pi}\sin 3t$.

11. The trigonometric identity $\cos 2t = 1 - 2\sin^2 t$ shows that $.\sin^2 t = \frac{1}{2} - \frac{1}{2}\cos 2t$. The expression on the right is in the subspace spanned by the trigonometric polynomials of order 3 or less, so this expression is the third-order Fourier approximation to $\sin^2 t$.

12. The trigonometric identity $\cos 3t = 4\cos^3 t - 3\cos t$ shows that $\cos^3 t = \frac{3}{4}\cos t + \frac{1}{4}\cos 3t$. The expression on the right is in the subspace spanned by the trigonometric polynomials of order 3 or less, so this expression is the third-order Fourier approximation to $\cos^3 t$.

13. Let f and g be in $C[0, 2\pi]$ and let m be a nonnegative integer. Then the linearity of the inner product shows that $\langle (f+g), \cos mt\rangle = \langle f, \cos mt\rangle + \langle g, \cos mt\rangle$ and $\langle (f+g), \sin mt\rangle = \langle f, \sin mt\rangle + \langle g, \sin mt\rangle$.

Dividing these identities respectively by $\langle \cos mt, \cos mt\rangle$ and $\langle \sin mt, \sin mt\rangle$ shows that the Fourier coefficients a_m and b_m for $f + g$ are the sums of the corresponding Fourier coefficients of f and of g.

14. Note that g and h are both in the subspace H spanned by the trigonometric polynomials of order 2 or less. Since h is the second-order Fourier approximation to f, it is closer to f than any other function in the subspace H.

15. **[M]** The weighting matrix W is the 13×13 diagonal matrix with diagonal entries 1, 1, 1, .9, .9, .8, .7, .6, .5, .4, .3, .2, .1. The design matrix X, parameter vector β, and observation vector $\mathbf{y}$ are:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 2^2 & 2^3 \\ 1 & 3 & 3^2 & 3^3 \\ 1 & 4 & 4^2 & 4^3 \\ 1 & 5 & 5^2 & 5^3 \\ 1 & 6 & 6^2 & 6^3 \\ 1 & 7 & 7^2 & 7^3 \\ 1 & 8 & 8^2 & 8^3 \\ 1 & 9 & 9^2 & 9^3 \\ 1 & 10 & 10^2 & 10^3 \\ 1 & 11 & 11^2 & 11^3 \\ 1 & 12 & 12^2 & 12^3 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0.0 \\ 8.8 \\ 29.9 \\ 62.0 \\ 104.7 \\ 159.1 \\ 222.0 \\ 294.5 \\ 380.4 \\ 471.1 \\ 571.7 \\ 686.8 \\ 809.2 \end{bmatrix}.$$

The design matrix X and the observation vector $\mathbf{y}$ are scaled by W:

$$WX = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 1.0 & 1.0 \\ 1.0 & 2.0 & 4.0 & 8.0 \\ .9 & 2.7 & 8.1 & 24.3 \\ .9 & 3.6 & 14.4 & 57.6 \\ .8 & 4.0 & 20.0 & 100.0 \\ .7 & 4.2 & 25.2 & 151.2 \\ .6 & 4.2 & 29.4 & 205.8 \\ .5 & 4.0 & 32.0 & 256.0 \\ .4 & 3.6 & 32.4 & 291.6 \\ .3 & 3.0 & 30.0 & 300.0 \\ .2 & 2.2 & 24.2 & 266.2 \\ .1 & 1.2 & 14.4 & 172.8 \end{bmatrix}, W\mathbf{y} = \begin{bmatrix} 0.00 \\ 8.80 \\ 29.90 \\ 55.80 \\ 94.23 \\ 127.28 \\ 155.40 \\ 176.70 \\ 190.20 \\ 188.44 \\ 171.51 \\ 137.36 \\ 80.92 \end{bmatrix}$$

Further compute

$$(WX)^T WX = \begin{bmatrix} 6.66 & 22.23 & 120.77 & 797.19 \\ 22.23 & 120.77 & 797.19 & 5956.13 \\ 120.77 & 797.19 & 5956.13 & 48490.23 \\ 797.19 & 5956.13 & 48490.23 & 420477.17 \end{bmatrix}, (WX)^T W\mathbf{y} = \begin{bmatrix} 747.844 \\ 4815.438 \\ 35420.468 \\ 285262.440 \end{bmatrix}$$

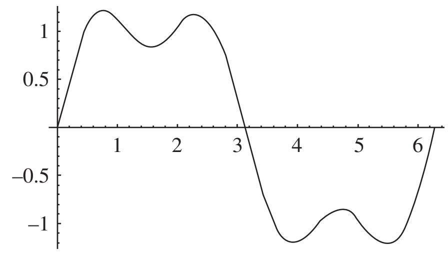
and find that $\hat{\boldsymbol{\beta}} = ((WX)^T WX)^{-1}(WX)^T W\mathbf{y} = \begin{bmatrix} -0.2685 \\ 3.6095 \\ 5.8576 \\ -0.0477 \end{bmatrix}$.

Thus the weighted least-squares cubic is $y = g(t) = -.2685 + 3.6095t + 5.8576t^2 - .0477t^3$. The velocity at $t = 4.5$ seconds is $g'(4.5) = 53.4$ ft./sec. This is about 0.7% faster than the estimate obtained in Exercise 13 of Section 6.6.

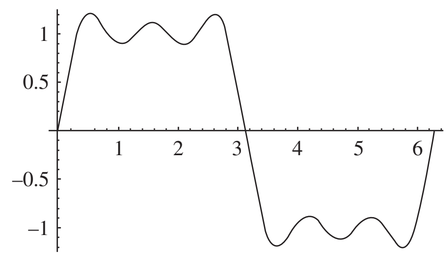
16. **[M]** Let $f(t) = \begin{cases} 1 & \text{for } 0 \le t < \pi \\ -1 & \text{for } \pi \le t < 2\pi \end{cases}$. The Fourier coefficients for f have already been found to be

$a_k = 0$ for all $k \ge 0$ and $b_k = \begin{cases} 4/(k\pi) & \text{for } k \text{ odd} \\ 0 & \text{for } k \text{ even} \end{cases}$. Thus

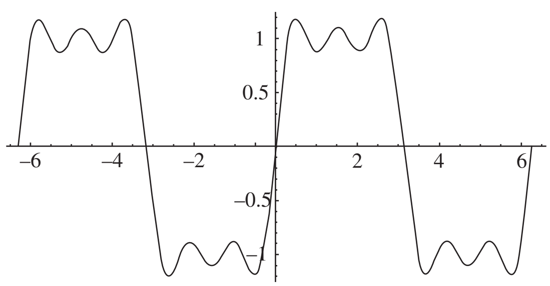
$f_4(t) = \dfrac{4}{\pi}\sin t + \dfrac{4}{3\pi}\sin 3t$ and $f_5(t) = \dfrac{4}{\pi}\sin t + \dfrac{4}{3\pi}\sin 3t + \dfrac{4}{5\pi}\sin 5t$. A graph of f_4 over the interval $[0, 2\pi]$ is



A graph of f_5 over the interval $[0, 2\pi]$ is



A graph of f_5 over the interval $[-2\pi, 2\pi]$ is



Chapter 6 SUPPLEMENTARY EXERCISES

1. **a**. False. The length of the zero vector is zero.

b. True. By the displayed equation before Example 2 in Section 6.1, with $c = -1$,
$\|-\mathbf{x}\| = \|(-1)\mathbf{x}\| = |-1|\,\|\mathbf{x}\| = \|\mathbf{x}\|$.

c. True. This is the definition of distance.

d. False. This equation would be true if $r\|\mathbf{v}\|$ were replaced by $|r|\,\|\mathbf{v}\|$.

e. False. Orthogonal *nonzero* vectors are linearly independent.

f. True. If $\mathbf{x}\cdot\mathbf{u} = 0$ and $\mathbf{x}\cdot\mathbf{v} = 0$, then $\mathbf{x}\cdot(\mathbf{u}-\mathbf{v}) = \mathbf{x}\cdot\mathbf{u} - \mathbf{x}\cdot\mathbf{v} = 0$.

g. True. This is the "only if" part of the Pythagorean Theorem in Section 6.1.

h. True. This is the "only if" part of the Pythagorean Theorem in Section 6.1 where $\mathbf{v}$ is replaced by $-\mathbf{v}$, because $\|-\mathbf{v}\|^2$ is the same as $\|\mathbf{v}\|^2$.

i. False. The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is a scalar multiple of $\mathbf{u}$, not $\mathbf{y}$ (except when $\mathbf{y}$ itself is already a multiple of $\mathbf{u}$).

j. True. The orthogonal projection of any vector $\mathbf{y}$ onto W is always a vector in W.

k. True. This is a special case of the statement in the box following Example 6 in Section 6.1 (and proved in Exercise 30 of Section 6.1).

l. False. The zero vector is in both W and $W^{\perp}$.

m. True. See Exercise 32 in Section 6.2. If $\mathbf{v}_i\cdot\mathbf{v}_j = 0$, then $(c_i\mathbf{v}_i)\cdot(c_j\mathbf{v}_j) = c_ic_j(\mathbf{v}_i\cdot\mathbf{v}_j) = c_ic_j0 = 0$.

n. False. This statement is true only for a *square* matrix. See Theorem 10 in Section 6.3.

o. False. An orthogonal matrix is square and has *orthonormal* columns.

p. True. See Exercises 27 and 28 in Section 6.2. If U has orthonormal columns, then $U^TU = I$. If U is also square, then the Invertible Matrix Theorem shows that U is invertible and $U^{-1} = U^T$. In this case, $UU^T = I$, which shows that the columns of U^T are orthonormal; that is, the rows of U are orthonormal.

q. True. By the Orthogonal Decomposition Theorem, the vectors $\text{proj}_W\,\mathbf{v}$ and $\mathbf{v} - \text{proj}_W\,\mathbf{v}$ are orthogonal, so the stated equality follows from the Pythagorean Theorem.

r. False. A least-squares solution is a vector $\hat{\mathbf{x}}$ (not $A\hat{\mathbf{x}}$) such that $A\hat{\mathbf{x}}$ is the closest point to $\mathbf{b}$ in Col A.

s. False. The equation $\hat{\mathbf{x}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b}$ describes the *solution* of the normal equations, not the matrix form of the normal equations. Furthermore, this equation makes sense only when A^TA is invertible.

2. If $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthonormal set and $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, then the vectors $c_1\mathbf{v}_1$ and $c_2\mathbf{v}_2$ are orthogonal (Exercise 32 in Section 6.2). By the Pythagorean Theorem and properties of the norm

$$\|\mathbf{x}\|^2 = \|c_1\mathbf{v}_1 + c_2\mathbf{v}_2\|^2 = \|c_1\mathbf{v}_1\|^2 + \|c_2\mathbf{v}_2\|^2 = (c_1\|\mathbf{v}_1\|)^2 + (c_2\|\mathbf{v}_2\|)^2 = |c_1|^2 + |c_2|^2$$

So the stated equality holds for $p = 2$. Now suppose the equality holds for $p = k$, with $k \ge 2$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ be an orthonormal set, and consider $\mathbf{x} = c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k + c_{k+1}\mathbf{v}_{k+1} = \mathbf{u}_k + c_{k+1}\mathbf{v}_{k+1}$, where $\mathbf{u}_k = c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k$. Observe that $\mathbf{u}_k$ and $c_{k+1}\mathbf{v}_{k+1}$ are orthogonal because $\mathbf{v}_j \cdot \mathbf{v}_{k+1} = 0$ for $j = 1,\ldots,k$. By the Pythagorean Theorem and the assumption that the stated equality holds for k, and because $\| c_{k+1}\mathbf{v}_{k+1} \|^2 = | c_{k+1} |^2 \| \mathbf{v}_{k+1} \|^2 = | c_{k+1} |^2$,

$$\| \mathbf{x} \|^2 = \| \mathbf{u}_k + c_{k+1}\mathbf{v}_{k+1} \|^2 = \| \mathbf{u}_k \|^2 + \| c_{k+1}\mathbf{v}_{k+1} \|^2 = | c_1 |^2 + \ldots + | c_{k+1} |^2$$

Thus the truth of the equality for $p = k$ implies its truth for $p = k + 1$. By the principle of induction, the equality is true for all integers $p \geq 2$.

3. Given $\mathbf{x}$ and an orthonormal set $\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}$ in $\mathbb{R}^n$, let $\hat{\mathbf{x}}$ be the orthogonal projection of $\mathbf{x}$ onto the subspace spanned by $\mathbf{v}_1,\ldots,\mathbf{v}_p$. By Theorem 10 in Section 6.3, $\hat{\mathbf{x}} = (\mathbf{x}\cdot\mathbf{v}_1)\mathbf{v}_1 + \ldots + (\mathbf{x}\cdot\mathbf{v}_p)\mathbf{v}_p$. By Exercise 2, $\| \hat{\mathbf{x}} \|^2 = | \mathbf{x}\cdot\mathbf{v}_1 |^2 + \ldots + | \mathbf{x}\cdot\mathbf{v}_p |^2$. Bessel's inequality follows from the fact that $\| \hat{\mathbf{x}} \|^2 \leq \| \mathbf{x} \|^2$, which is noted before the proof of the Cauchy-Schwarz inequality in Section 6.7.

4. By parts (a) and (c) of Theorem 7 in Section 6.2, $\{U\mathbf{v}_1,\ldots,U\mathbf{v}_k\}$ is an orthonormal set in $\mathbb{R}^n$. Since there are n vectors in this linearly independent set, the set is a basis for $\mathbb{R}^n$.

5. Suppose that $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ for all $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^n$, and let $\mathbf{e}_1,\ldots,\mathbf{e}_n$ be the standard basis for $\mathbb{R}^n$. For $j = 1, \ldots, n$, $U\mathbf{e}_j$ is the jth column of U. Since $\| U\mathbf{e}_j \|^2 = (U\mathbf{e}_j)\cdot(U\mathbf{e}_j) = \mathbf{e}_j\cdot\mathbf{e}_j = 1$, the columns of U are unit vectors; since $(U\mathbf{e}_j)\cdot(U\mathbf{e}_k) = \mathbf{e}_j\cdot\mathbf{e}_k = 0$ for $j \neq k$, the columns are pairwise orthogonal.

6. If $U\mathbf{x} = \lambda\mathbf{x}$ for some $\mathbf{x} \neq \mathbf{0}$, then by Theorem 7(a) in Section 6.2 and by a property of the norm, $\| \mathbf{x} \| = \| U\mathbf{x} \| = \| \lambda\mathbf{x} \| = | \lambda | \| \mathbf{x} \|$, which shows that $| \lambda | = 1$, because $\mathbf{x} \neq \mathbf{0}$.

7. Let $\mathbf{u}$ be a unit vector, and let $Q = I - 2\mathbf{u}\mathbf{u}^T$. Since $(\mathbf{u}\mathbf{u}^T)^T = \mathbf{u}^{TT}\mathbf{u}^T = \mathbf{u}\mathbf{u}^T$,

$$Q^T = (I - 2\mathbf{u}\mathbf{u}^T)^T = I - 2(\mathbf{u}\mathbf{u}^T)^T = I - 2\mathbf{u}\mathbf{u}^T = Q$$

Then

$$QQ^T = Q^2 = (I - 2\mathbf{u}\mathbf{u}^T)^2 = I - 2\mathbf{u}\mathbf{u}^T - 2\mathbf{u}\mathbf{u}^T + 4(\mathbf{u}\mathbf{u}^T)(\mathbf{u}\mathbf{u}^T)$$

Since $\mathbf{u}$ is a unit vector, $\mathbf{u}^T\mathbf{u} = \mathbf{u}\cdot\mathbf{u} = 1$, so $(\mathbf{u}\mathbf{u}^T)(\mathbf{u}\mathbf{u}^T) = \mathbf{u}(\mathbf{u}^T)(\mathbf{u})\mathbf{u}^T = \mathbf{u}\mathbf{u}^T$, and

$$QQ^T = I - 2\mathbf{u}\mathbf{u}^T - 2\mathbf{u}\mathbf{u}^T + 4\mathbf{u}\mathbf{u}^T = I$$

Thus Q is an orthogonal matrix.

8. **a**. Suppose that $\mathbf{x}\cdot\mathbf{y} = 0$. By the Pythagorean Theorem, $\| \mathbf{x} \|^2 + \| \mathbf{y} \|^2 = \| \mathbf{x} + \mathbf{y} \|^2$. Since T preserves lengths and is linear,

$$\| T(\mathbf{x}) \|^2 + \| T(\mathbf{y}) \|^2 = \| T(\mathbf{x} + \mathbf{y}) \|^2 = \| T(\mathbf{x}) + T(\mathbf{y}) \|^2$$

This equation shows that $T(\mathbf{x})$ and $T(\mathbf{y})$ are orthogonal, because of the Pythagorean Theorem. Thus T preserves orthogonality.

b. The standard matrix of T is $\begin{bmatrix} T(\mathbf{e}_1) & \ldots & T(\mathbf{e}_n) \end{bmatrix}$, where $\mathbf{e}_1,\ldots,\mathbf{e}_n$ are the columns of the identity matrix. Then $\{T(\mathbf{e}_1),\ldots,T(\mathbf{e}_n)\}$ is an orthonormal set because T preserves both orthogonality and lengths (and because the columns of the identity matrix form an orthonormal set). Finally, a square matrix with orthonormal columns is an orthogonal matrix, as was observed in Section 6.2.

9. Let $W = \text{Span}\{\mathbf{u}, \mathbf{v}\}$. Given $\mathbf{z}$ in $\mathbb{R}^n$, let $\hat{\mathbf{z}} = \text{proj}_W \mathbf{z}$. Then $\hat{\mathbf{z}}$ is in Col A, where $A = [\mathbf{u} \quad \mathbf{v}]$. Thus there is a vector, say, $\hat{\mathbf{x}}$ in $\mathbb{R}^2$, with $A\hat{\mathbf{x}} = \hat{\mathbf{z}}$. So, $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{z}$. The normal equations may be solved to find $\hat{\mathbf{x}}$, and then $\hat{\mathbf{z}}$ may be found by computing $A\hat{\mathbf{x}}$.

10. Use Theorem 14 in Section 6.5. If $c \neq 0$, the least-squares solution of $A\mathbf{x} = c\mathbf{b}$ is given by $(A^TA)^{-1}A^T(c\mathbf{b})$, which equals $c(A^TA)^{-1}A^T\mathbf{b}$, by linearity of matrix multiplication. This solution is c times the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

11. Let $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix}$, and $A = \begin{bmatrix} \mathbf{v}^T \\ \mathbf{v}^T \\ \mathbf{v}^T \end{bmatrix} = \begin{bmatrix} 1 & -2 & 5 \\ 1 & -2 & 5 \\ 1 & -2 & 5 \end{bmatrix}$. Then the given set of equations is $A\mathbf{x} = \mathbf{b}$, and the set of all least-squares solutions coincides with the set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$. The column-row expansions of A^TA and $A^T\mathbf{b}$ give

$$A^TA = \mathbf{v}\mathbf{v}^T + \mathbf{v}\mathbf{v}^T + \mathbf{v}\mathbf{v}^T = 3\mathbf{v}\mathbf{v}^T, A^T\mathbf{b} = a\mathbf{v} + b\mathbf{v} + c\mathbf{v} = (a+b+c)\mathbf{v}$$

Thus $A^TA\mathbf{x} = 3(\mathbf{v}\mathbf{v}^T)\mathbf{x} = 3\mathbf{v}(\mathbf{v}^T\mathbf{x}) = 3(\mathbf{v}^T\mathbf{x})\mathbf{v}$ since $\mathbf{v}^T\mathbf{x}$ is a scalar, and the normal equations have become $3(\mathbf{v}^T\mathbf{x})\mathbf{v} = (a+b+c)\mathbf{v}$, so $3(\mathbf{v}^T\mathbf{x}) = a+b+c$, or $\mathbf{v}^T\mathbf{x} = (a+b+c)/3$. Computing $\mathbf{v}^T\mathbf{x}$ gives the equation $x - 2y + 5z = (a + b + c)/3$ which must be satisfied by all least-squares solutions to $A\mathbf{x} = \mathbf{b}$.

12. The equation (1) in the exercise has been written as $V\lambda = \mathbf{b}$, where V is a single nonzero column vector $\mathbf{v}$, and $\mathbf{b} = A\mathbf{v}$. The least-squares solution $\hat{\lambda}$ of $V\lambda = \mathbf{b}$ is the exact solution of the normal equations $V^TV\lambda = V^T\mathbf{b}$. In the original notation, this equation is $\mathbf{v}^T\mathbf{v}\lambda = \mathbf{v}^TA\mathbf{v}$. Since $\mathbf{v}^T\mathbf{v}$ is nonzero, the least squares solution $\hat{\lambda}$ is $\mathbf{v}^TA\mathbf{v}/(\mathbf{v}^T\mathbf{v})$. This expression is the Rayleigh quotient discussed in the Exercises for Section 5.8.

13. **a**. The row-column calculation of $A\mathbf{u}$ shows that each row of A is orthogonal to every $\mathbf{u}$ in Nul A. So each row of A is in $(\text{Nul } A)^\perp$. Since $(\text{Nul } A)^\perp$ is a subspace, it must contain all linear combinations of the rows of A; hence $(\text{Nul } A)^\perp$ contains Row A.

b. If rank $A = r$, then dimNul $A = n - r$ by the Rank Theorem. By Exercsie 24(c) in Section 6.3, $\text{dimNul } A + \dim(\text{Nul } A)^\perp = n$, so $\dim(\text{Nul } A)^\perp$ must be r. But Row A is an r-dimensional subspace of $(\text{Nul } A)^\perp$ by the Rank Theorem and part (a). Therefore, $\text{Row } A = (\text{Nul } A)^\perp$.

c. Replace A by A^T in part (b) and conclude that $\text{Row } A^T = (\text{Nul } A^T)^\perp$. Since $\text{Row } A^T = \text{Col } A$, $\text{Col } A = (\text{Nul } A^T)^\perp$.

14. The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is in Col A. By Exercise 13(c), $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is orthogonal to $\text{Nul } A^T$. This happens if and only if $\mathbf{b}$ is orthogonal to all solutions of $A^T\mathbf{x} = \mathbf{0}$.

15. If $A = URU^T$ with U orthogonal, then A is similar to R (because U is invertible and $U^T = U^{-1}$), so A has the same eigenvalues as R by Theorem 4 in Section 5.2. Since the eigenvalues of R are its n real diagonal entries, A has n real eigenvalues.

16. **a**. If $U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n]$, then $AU = [\lambda_1\mathbf{u}_1 \quad A\mathbf{u}_2 \quad \dots \quad A\mathbf{u}_n]$. Since $\mathbf{u}_1$ is a unit vector and $\mathbf{u}_2, \dots, \mathbf{u}_n$ are orthogonal to $\mathbf{u}_1$, the first column of $U^T AU$ is $U^T(\lambda_1\mathbf{u}_1) = \lambda_1 U^T\mathbf{u}_1 = \lambda_1\mathbf{e}_1$.

b. From (a),

$$U^T AU = \begin{bmatrix} \lambda_1 & * & * & * & * \\ 0 & & & & \\ \vdots & & A_1 & & \\ 0 & & & & \end{bmatrix}$$

View $U^T AU$ as a 2×2 block upper triangular matrix, with A_1 as the (2, 2)-block. Then from Supplementary Exercise 12 in Chapter 5,

$$\det(U^T AU - \lambda I_n) = \det(\lambda_1 - \lambda I_1) \cdot \det(A_1 - \lambda I_{n-1}) = (\lambda_1 - \lambda) \cdot \det(A_1 - \lambda I_{n-1})$$

This shows that the eigenvalues of $U^T AU$, namely, $\lambda_1, \dots, \lambda_n$, consist of λ_1 and the eigenvalues of A_1. So the eigenvalues of A_1 are $\lambda_2, \dots, \lambda_n$.

17. **[M]** Compute that $\|\Delta\mathbf{x}\|/\|\mathbf{x}\| = .4618$ and $\text{cond}(A) \times (\|\Delta\mathbf{b}\|/\|\mathbf{b}\|) = 3363 \times (1.548 \times 10^{-4}) = .5206$. In this case, $\|\Delta\mathbf{x}\|/\|\mathbf{x}\|$ is almost the same as $\text{cond}(A) \times \|\Delta\boldsymbol{b}\|/\|\mathbf{b}\|$.

18. **[M]** Compute that $\|\Delta\mathbf{x}\|/\|\mathbf{x}\| = .00212$ and $\text{cond}(A) \times (\|\Delta\mathbf{b}\|/\|\mathbf{b}\|) = 3363 \times (.00212) \approx 7.130$. In this case, $\|\Delta\mathbf{x}\|/\|\mathbf{x}\|$ is almost the same as $\|\Delta\mathbf{b}\|/\|\mathbf{b}\|$, even though the large condition number suggests that $\|\Delta\mathbf{x}\|/\|\mathbf{x}\|$ could be much larger.

19. **[M]** Compute that $\|\Delta\mathbf{x}\| / \|\mathbf{x}\| = 7.178 \times 10^{-8}$ and $\text{cond}(A) \times (\|\Delta\mathbf{b}\| / \|\mathbf{b}\|) = 23683 \times (2.832 \times 10^{-4}) =$ 6.707. Observe that the relative change in $\mathbf{x}$ is *much* smaller than the relative change in $\mathbf{b}$. In fact the theoretical bound on the relative change in $\mathbf{x}$ is 6.707 (to four significant figures). This exercise shows that even when a condition number is large, the relative error in the solution need not be as large as you suspect.

20. **[M]** Compute that $\|\Delta\mathbf{x}\|/\|\mathbf{x}\| = .2597$ and $\text{cond}(A) \times (\|\Delta\mathbf{b}\| / \|\mathbf{b}\|) = 23683 \times (1.097 \times 10^{-5}) = .2598$. This calculation shows that the relative change in $\mathbf{x}$, for this particular $\mathbf{b}$ and $\Delta\mathbf{b}$, should not exceed .2598. In this case, the theoretical maximum change is almost acheived.